Assessing the quality of convex approximations for two-stage totally unimodular integer recourse models

Romeijnders, Ward; Morton, D.; van der Vlerk, Maarten

Published in:
INFORMS Journal on Computing

DOI:
10.1287/ijoc.2016.0725

Publication date:
2017

Submitted to *INFORMS Journal on Computing*
manuscript (Please, provide the mansucript number!)

Assessing the quality of convex approximations for two-stage totally unimodular integer recourse models

Ward Romeijnders
Department of Operations, University of Groningen, P.O. Box 800, 9700 AV, Groningen, The Netherlands, w.romeijnders@rug.nl

David P. Morton
Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA, david.morton@northwestern.edu

Maarten H. van der Vlerk
Department of Operations, University of Groningen, P.O. Box 800, 9700 AV, Groningen, The Netherlands, m.h.van.der.vlerk@rug.nl

We consider two types of convex approximations of two-stage totally unimodular integer recourse models. Although worst-case error bounds are available for these approximations, their actual performance has not yet been investigated, mainly because this requires solving the original recourse model. In this paper we assess the quality of the approximating solutions using Monte Carlo sampling, or more specifically, using the so-called multiple replications procedure. Based on numerical experiments for an integer newsvendor problem, a fleet allocation and routing problem, and a stochastic activity network investment problem, we conclude that the error bounds are reasonably sharp if the variability of the random parameters in the model is either small or large; otherwise, the actual error of using the convex approximations is much smaller than the error bounds suggests. Moreover, we conclude that the solutions obtained using the convex approximations are good only if the variability of the random parameters is medium to large. In case this variability is small, however, typically sampling methods perform best, even with modest sample sizes. In this sense, the convex approximations and sampling methods can be considered as complementary solution methods. Moreover, as required for our applications, we extend our approach to derive new error bounds dealing with deterministic second-stage side constraints and relatively complete recourse, and perfect dependencies in the right-hand side vector.

Key words: Stochastic Programming, Integer Recourse, Convex Approximations, Sampling Methods

1. Introduction

We consider the stochastic optimization problem

$$\eta^* := \min_x \Big\{ \mathbb{E}_\omega[g(x,\omega)] : x \in X \Big\}, \tag{1}$$

where the uncertainty is explicitly modeled using the random vector ω (with known distribution function F) and the objective is to find optimal here-and-now decisions $x \in X$ to minimize the expected value function $G(x) := \mathbb{E}_\omega[g(x,\omega)]$.

In its general form, model (1) can represent many stochastic programming problems (see, e.g., Birge and Louveaux (2011), Prékopa (1995), Shapiro et al. (2009), Wallace and Ziemba (2005)). However, throughout this paper we restrict attention to two-stage integer recourse models, where $X \subset \mathbb{R}^{n_1}_+$, $Y \subset \mathbb{R}^{n_2}_+$ are polyhedral sets, and g is defined for every $x \in X$ and $\omega \in \Omega$ as

$$g(x,\omega) = cx + \min_y \{ q(\omega)y : Wy \geq \zeta(\omega) - T(\omega)x,\ y \in Y \cap \mathbb{Z}^{n_2} \}. \tag{2}$$

Here, the cost vector $q(\omega)$, right-hand side vector $\zeta(\omega)$, and technology matrix $T(\omega)$ are random and depend on the underlying random vector ω. We introduce the notation in (1) since several of our ideas, methods, and results also hold in this more general setting.

The function g in (2) is called an integer recourse function because of the integer-constrained recourse variables y. Such decision variables arise often in practice to model indivisibilities or on-off decisions. With the corresponding model (1) in mind, these recourse variables y are determined after realization of the random vector ω, and can be used to compensate for infeasibilities of the underlying random goal constraints $T(\omega)x \geq \zeta(\omega)$.

The integer recourse function g is generally non-convex in x for every $\omega \in \Omega$ because of the integer programming problem involved. As a consequence, the expected value function G is generally non-convex as well (Rinnooy Kan and Stougie 1988); see Klein Haneveld et al. (2006) for exceptions in the simple integer recourse case. This lack of convexity is the main reason why integer recourse models are much harder to solve than their continuous counterparts. Indeed, for the latter type of problems, efficient algorithms such as the L-shaped method (van Slyke and Wets 1969), regularized decomposition (Ruszczyński 1986), and stochastic decomposition (Higle and Sen 1991) are available that explicitly use convexity (see Zverovich et al. (2012) for a numerical study comparing several algorithms).

A possible approach for solving integer recourse models is to replace g in (1) by a function $\hat{g} : X \times \Omega \mapsto \mathbb{R}$ that is *convex* in x for every $\omega \in \Omega$. Then, the approximating model

$$\hat{\eta} := \min_x \{\hat{G}(x) : x \in X\} \tag{3}$$

with $\hat{G}(x) := \mathbb{E}_\omega[\hat{g}(x,\omega)]$, $x \in X$, can be solved using tools from convex optimization (yielding an approximating solution $\hat{x}$). For the special case of totally unimodular (TU) integer recourse models, i.e., for TU recourse matrices W, with $Y = \mathbb{R}^{n_2}_+$ and deterministic q and T, such convex approximations have been developed (van der Vlerk 2004, Romeijnders et al. 2016). In fact, these references show that model (3) corresponding to these approximations can be represented as a continuous recourse model and can thus be solved efficiently using one of the methods mentioned above. A question that remains, one of the main topics of this paper, concerns the quality of the approximating solution, $\hat{x}$, of model (3).

Romeijnders et al. (2015, 2016) measure the performance of the convex approximations by upper bounds $U(G,\hat{G})$ on $\|G - \hat{G}\|_\infty := \sup\{|G(x) - \hat{G}(x)| : x \in X\}$. Such upper bounds are useful since

$$|\hat{\eta} - \eta^*| \leq \|G - \hat{G}\|_\infty \leq U(G, \hat{G})$$

and

$$G(\hat{x}) - \eta^* \leq 2\|G - \hat{G}\|_\infty \leq 2U(G, \hat{G}). \tag{4}$$

See, e.g., Romeijnders et al. (2016) for a proof of these results for the TU integer recourse case. Numerical experiments in Romeijnders et al. (2015) for several (small) examples suggest that the second inequality in (4) is reasonably tight if the upper bounds $U(G,\hat{G})$ of Romeijnders et al. (2015, 2016) are used. The sharpness of the first inequality in (4), however, has not yet been investigated, and doing so is another main topic of this paper. The central difficulty in assessing the sharpness of the bounds, and in fact the motivation for deriving convex approximations of g, is that it is very hard to solve the original integer recourse model (1) to obtain η^*, especially for larger problem instances.

In this paper we assess the quality of $\hat{x}$, and the sharpness of the first inequality in (4), using sampling. In particular, we will use the multiple replications procedure (MRP) developed in Mak et al. (1999). We carry out numerical experiments for an integer newsvendor problem, for a fleet allocation and routing problem, and for an investment problem on a stochastic activity network. These experiments show that the solutions obtained using the convex approximations are good if the 'variability' of the random parameters in the models is medium to large. In addition, in case of medium 'variability' the performance of the convex approximations is much better than their error bounds suggest; i.e., in this case the error bounds are not so sharp. On the other hand, if this 'variability' is small then the solutions obtained using the convex approximations are not so good. However, these are precisely the cases in which sampling methods can work quite well with modest sample sizes, and in this sense we may view convex approximations and sampling methods as complementary approaches for approximately solving TU integer recourse models.

Summarizing, the contribution of this paper is threefold.

(i) We evaluate the sharpness of existing (and new) error bounds on the optimality gap, $G(\hat{x}) - \eta^*$, of optimal solutions to convex approximations, $\hat{x}$.

(ii) We assess the quality of approximating solutions $\hat{x}$ and x^S, obtained from convex approximations and sampling methods, respectively.

(iii) We compare the relative performance of solutions obtained from convex approximations and sampling methods.

The remainder of this paper is organized as follows. In Section 2 we discuss the literature on convex approximations for integer recourse models and the literature on sampling methods for assessing the quality of candidate solutions in stochastic programs. In Sections 3–5 we show numerical experiments for an integer newsvendor problem, a fleet allocation and routing problem, and a stochastic activity network investment problem, respectively. For the latter two problems we have to extend the analysis of Romeijnders et al. (2016) to derive an error bound for the convex approximation to deal with deterministic second-stage side constraints and relatively complete recourse instead of complete recourse (Section 4), and to deal with perfect dependencies in the right-hand side random vector (Section 5). Finally, Section 6 comprises a summary and conclusions.

2. Literature review

We review the literature on both solution methods for integer recourse models (Section 2.1) and sampling methods for assessing the quality of candidate solutions in stochastic programming problems (Section 2.2). Our focus is on convex approximations of integer recourse models, their error bounds, and the multiple replications procedure (MRP) to be used in Sections 3–5.

2.1. Solution methods for integer recourse models

During the last decades a variety of solution methods have been developed for integer recourse models, including the integer L-shaped method (Laporte and Louveaux 1993), dual decomposition (Carøe and Schultz 1999), branch-and-bound (Ahmed et al. 2004), and disjunctive decomposition (Sen and Higle 2005). These solution methods typically combine ideas from deterministic integer programming and stochastic continuous programming, and are aimed at finding (near-)optimal solutions. This is the main reason why these methods, in general, have difficulties solving very large problem instances, motivating the development of convex approximations.

In the remainder of this subsection we restrict our attention to the literature on these convex approximations and their error bounds. For readers interested in other solution methods for integer recourse models we refer to the survey papers of Klein Haneveld and van der Vlerk (1999), Louveaux and Schultz (2003), Schultz (2003), and Sen (2005).

2.1.1. Convex approximations for integer recourse models. Klein Haneveld et al. (2006) were the first to develop a class of convex approximations for the special case of simple integer recourse models. These so-called *α-approximations* were later extended by van der Vlerk to the cases of TU integer recourse (van der Vlerk 2004) and simple mixed-integer recourse (van der Vlerk 2010). The recurring idea in these approximations (see also Section 3 of the survey paper of Romeijnders et al. (2014)) is to simultaneously relax the integrality constraints in the model defining g and perturb the distribution of the random right-hand side ω. For g defined in (2) with $Y = \mathbb{R}^{n_2}_+$, $\zeta(\omega) = \omega$, and deterministic q and T, this yields, for every $\alpha \in \mathbb{R}^m$,

$$g_\alpha(x,\omega) := cx + \min_y \left\{ qy : Wy \geq \lceil \omega \rceil_\alpha - Tx,\; y \in \mathbb{R}^{n_2}_+ \right\}, \quad x \in X,\, \omega \in \Omega, \tag{5}$$

where $\lceil \omega \rceil_\alpha := \lceil \omega - \alpha \rceil + \alpha$ is a discrete random vector with support contained in $\alpha + \mathbb{Z}^m$. As already mentioned in the introduction, the resulting approximating problem with g replaced by g_α in (1) corresponds to a (convex) continuous recourse model with a discrete distribution that, with existing algorithms, is more computationally tractable.

An alternative convex approximation that also can be represented as a continuous recourse model is the so-called *shifted LP-relaxation approximation* developed by Romeijnders et al. (2016), defined as

$$\hat{g}(x,\omega) = cx + \min_y \left\{ qy : Wy \geq \omega + \frac{1}{2}e_m - Tx,\; y \in \mathbb{R}^{n_2}_+ \right\}, \quad x \in X,\, \omega \in \Omega, \tag{6}$$

where e_m is the m-dimensional all-one vector. The error bound of this approximation (see Theorem 1 below) improves the bound of the α-approximation by a factor 2. Moreover, the bound is tight in a worst-case sense (Romeijnders et al. 2016).

2.1.2. Error bounds for convex approximations of TU integer recourse models. Error bounds, i.e., upper bounds on $\|G - G_\alpha\|_\infty$ and $\|G - \hat{G}\|_\infty$ with $G_\alpha(x) := \mathbb{E}_\omega[g_\alpha(x,\omega)]$ and $\hat{G}(x) := \mathbb{E}_\omega[\hat{g}(x,\omega)]$, $x \in X$, are derived under several assumptions:

(A1) Complete recourse: $g(x,\omega) < +\infty$ for every $x \in \mathbb{R}^{n_1}$ and $\omega \in \mathbb{R}^m$.

(A2) Sufficiently expensive (or dual feasible) recourse: $\Lambda := \{\lambda \in \mathbb{R}^m_+ : \lambda W \leq q\} \neq \emptyset$.

(A3) Finite expectations: $\mathbb{E}_\omega[|\omega_i|] < +\infty$ for every $i = 1,\dots,m$.

In Section 4 we consider a problem where the complete recourse assumption is violated. Instead, a relaxation of this assumption holds:

(A1') Relatively complete recourse: $g(x,\omega) < +\infty$ for every $x \in X$ and $\omega \in \Omega$.

As we show in Section 4 this has consequences for the type of convex approximation to use and its corresponding error bound.

Theorem 1 below shows error bounds for α-approximations and the shifted LP-relaxation approximation. They correspond to upper bounds $U(G,G_\alpha)$ and $U(G,\hat{G})$ that can appear on the right-hand side of (4), and in Sections 3–5 we compare them with the optimality gaps $G(x_\alpha) - \eta^*$ and $G(\hat{x}) - \eta^*$. A detailed proof of Theorem 1 is omitted here and can be found in Romeijnders et al. (2016). We do discuss the main line of the proof in Section 2.1.3 because it facilitates our proofs for error bounds for two specific problems in Sections 4 and 5.

The error bounds for the approximations depend on the total variations of the probability density functions of the random variables in the model.

DEFINITION 1. Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a real-valued function, and let $I \subset \mathbb{R}$ be an interval. Let $\Pi(I)$ denote the set of all finite ordered sets $P = \{t_1,\dots,t_{N+1}\}$ with $t_1 < \cdots < t_{N+1}$ in I. Then, the *total variation* of f on I, denoted $|\Delta|f(I)$, is defined as

$$|\Delta|f(I) = \sup_{P \in \Pi(I)} V_f(P),$$

where

$$V_f(P) = \sum_{i=1}^{N} |f(t_{i+1}) - f(t_i)|.$$

We will write $|\Delta|f := |\Delta|f(\mathbb{R})$.

THEOREM 1. *Consider the totally unimodular integer recourse function*

$$g(x,\omega) := cx + \min_y \left\{ qy : Wy \geq \omega - Tx,\, y \in \mathbb{Z}_+^{n_2} \right\}, \quad x \in X,\, \omega \in \Omega,$$

where ω is a continuous random vector with joint pdf f and with independently distributed components. Let g_α and $\hat{g}$ denote the α-approximation and shifted LP-relaxation approximation defined in (5) and (6), respectively, with G_α and $\hat{G}$ denoting their expected value functions. Then, under assumptions (A1)-(A3) we have for every $\alpha \in \mathbb{R}^m$,

$$\|G - G_\alpha\|_\infty \leq \sum_{i=1}^m \lambda_i^* h(|\Delta| f_i) = U(G, G_\alpha)$$

and

$$\|G - \hat{G}\|_\infty \leq \frac{1}{2} \sum_{i=1}^m \lambda_i^* h(|\Delta| f_i) = U(G, \hat{G}),$$

where $\lambda_i^* := \max\{\lambda_i : \lambda W \leq q,\, \lambda \in \mathbb{R}_+^m\}$, $|\Delta| f_i$ denotes the total variation of the i-th marginal density function, f_i, and $h : (0,\infty) \mapsto \mathbb{R}$ is defined as

$$h(t) = \begin{cases} t/8, & t \leq 4 \\ 1 - 2/t, & t \geq 4. \end{cases} \tag{7}$$

REMARK 1. The assumption in Theorem 1 that the components of ω are independently distributed is not necessary. Indeed, in Section 5 we deal with a special type of perfect dependency in the right-hand side, and in Romeijnders et al. (2015, 2016) bounds for the dependent case are derived involving conditional density functions instead of marginal ones. The form in which we present Theorem 1 eases exposition given the numerical experiments we consider in Sections 3–5.

The error bounds in Theorem 1 are smaller if the total variations $|\Delta| f_i$ of the marginal densities f_i are smaller. For example, for a normally distributed random vector, ω, this implies that we expect the performance of the convex approximations to be better if the standard deviations are larger. We will confirm this conjecture by numerical experiments in Sections 3–5.

2.1.3. Main line of the proof of Theorem 1. Since the derivation of the error bounds in Theorem 1 is very similar for both the α-approximation and the shifted LP-relaxation approximation, we only discuss the proof for the α-approximation.

First, we derive a dual representation of the optimization problems in g and g_α. Since W is TU, we have for every $x \in X$ and $\omega \in \Omega$,

$$g(x,\omega) = cx + \min_y \Big\{ qy : Wy \geq \lceil \omega - Tx \rceil,\ y \in \mathbb{R}_+^{n_2} \Big\} \tag{8}$$

$$= cx + \max_\lambda \Big\{ \lambda \lceil \omega - Tx \rceil : \lambda \in \Lambda \Big\}, \tag{9}$$

where the second equality follows from strong LP duality and where the dual feasible region $\Lambda := \{\lambda \in \mathbb{R}_+^m : \lambda W \leq q\}$ is non-empty and bounded by assumptions (A1) and (A2). Similarly, for the α-approximation we have for every $\alpha \in \mathbb{R}^m$,

$$g_\alpha(x,\omega) = cx + \max_\lambda \Big\{ \lambda(\lceil \omega \rceil_\alpha - Tx) : \lambda \in \Lambda \Big\}, \quad x \in X, \omega \in \Omega. \tag{10}$$

Suppose for the moment that the dual feasible region Λ contains only a single point. Then, for every fixed $x \in X$, $\omega \in \Omega$, and defining tender variables $z := Tx$,

$$\begin{aligned} g(x,\omega) - g_\alpha(x,\omega) &= \lambda\Big(\lceil \omega - Tx \rceil + Tx - \lceil \omega \rceil_\alpha \Big) \\ &= \sum_{i=1}^m \lambda_i \Big(\lceil \omega_i - z_i \rceil + z_i - \lceil \omega_i \rceil_{\alpha_i} \Big) \\ &= \sum_{i=1}^m \lambda_i \Big(\lceil \omega_i \rceil_{z_i} - \lceil \omega_i \rceil_{\alpha_i} \Big). \end{aligned}$$

Thus, for fixed z and α the difference $g - g_\alpha$ can be decomposed componentwise in ω_i. Moreover, all properties of $g - g_\alpha$ follow directly from those of the one-dimensional function $\bar{\varphi}_{z_i,\alpha_i}$ given in Definition 2.

Definition 2. For every $z_i \in \mathbb{R}$ and $\alpha_i \in \mathbb{R}$ we define the function $\bar{\varphi}_{z_i,\alpha_i} : \mathbb{R} \mapsto \mathbb{R}$ as

$$\bar{\varphi}_{z_i,\alpha_i}(t) = \lceil t \rceil_{z_i} - \lceil t \rceil_{\alpha_i} = \Big(\lceil t - z_i \rceil + z_i \Big) - \Big(\lceil t - \alpha_i \rceil + \alpha_i \Big), \quad t \in \mathbb{R}.$$

Moreover, for every $z_i \in \mathbb{R}$ we define $\hat{\varphi}_{z_i} : \mathbb{R} \mapsto \mathbb{R}$ as

$$\hat{\varphi}_{z_i}(t) = \lceil t \rceil_{z_i} - \Big(t + 1/2 \Big) = \Big(\lceil t - z_i \rceil + z_i \Big) - \Big(t + 1/2 \Big), \quad t \in \mathbb{R}.$$

The function $\hat{\varphi}_{z_i}$ can be interpreted as the underlying difference function of the shifted LP-relaxation approximation, which we use in our numerical experiments. Both functions $\bar{\varphi}_{z_i,\alpha_i}$ and $\hat{\varphi}_{z_i}$ are periodic in t with period $p=1$ and mean value $p^{-1}\int_0^p \bar{\varphi}_{z_i,a_i}(t)dt = p^{-1}\int_0^p \hat{\varphi}_{z_i}(t)dt = 0$. We use these properties to bound $\mathbb{E}_{\omega_i}[\bar{\varphi}_{z_i,\alpha_i}(\omega_i)]$, yielding an upper bound on $\|G - G_\alpha\|_\infty$ since

$$G(x) - G_\alpha(x) = \sum_{i=1}^m \lambda_i \mathbb{E}_{\omega_i}\Big[\bar{\varphi}_{z_i,\alpha_i}(\omega_i)\Big].$$

Surprisingly, in the general case (without the assumption that Λ is a singleton) the analysis above is still helpful since it turns out that we are allowed to 'round up' the λ's to a single vector λ^* with $\lambda \leq \lambda^*$ for every $\lambda \in \Lambda$. Below we illustrate this idea by deriving an upper bound for $G(x) - G_\alpha(x)$, $x \in X$. A lower bound can be obtained in a similar way.

Let $x \in X$ be given and let $\lambda(\omega) \in \Lambda$ denote maximizers in model (9) for each $\omega \in \Omega$. Since $\lambda(\omega)$ is feasible but not necessarily optimal in model (10), we have

$$g(x,\omega) - g_\alpha(x,\omega) \leq \sum_{i=1}^m \lambda_i(\omega)\Big(\lceil \omega_i \rceil_{z_i} - \lceil \omega_i \rceil_{\alpha_i} \Big) = \sum_{i=1}^m \lambda_i(\omega)\bar{\varphi}_{z_i,\alpha_i}(\omega_i).$$

Moreover, it is not hard to show that $\lambda_i(\omega)$ is monotone non-decreasing in ω_i for every $\omega_{(i)} \in \mathbb{R}^{m-1}$, where $\omega_{(i)}$ denotes ω without its i-th component. This monotonicity property is one of the assumptions in Proposition 1, which is key to 'round up' $\lambda_i(\omega)$ to λ_i^*.

The proof of Proposition 1 can be found in Romeijnders et al. (2016). Observe that we use $\mathcal{F}$ to denote the set of probability density functions f of bounded variation (i.e., $|\Delta| f < +\infty$).

Proposition 1. *Let $\lambda : \mathbb{R} \mapsto \mathbb{R}$ be a real-valued monotone function such that $0 \leq \lambda(x) \leq \lambda^*$ for all $x \in \mathbb{R}$, and let $\varphi : \mathbb{R} \mapsto \mathbb{R}$ be a bounded periodic function with period p and mean value $p^{-1}\int_0^p \varphi(x)dx = 0$. Then, for every continuous random variable ω with probability density function $f \in \mathcal{F}$,*

$$-\lambda^* M(-\varphi, |\Delta| f) \leq \mathbb{E}_\omega[\lambda(\omega)\varphi(\omega)] \leq \lambda^* M(\varphi, |\Delta| f), \tag{11}$$

where for every $B \in \mathbb{R}$ with $B > 0$,

$$M(\varphi, B) := \sup_{f \in \mathcal{F}} \Big\{ \mathbb{E}_\omega[\varphi(\omega)] : |\Delta| f \leq B \Big\}. \tag{12}$$

Using Proposition 1 we are able to reduce the problem of finding an upper bound on $\|G - G_\alpha\|_\infty$ to the bound of (12) since for every $x \in X$ and $\alpha \in \mathbb{R}^m$,

$$G(x) - G_\alpha(x) \leq \sum_{i=1}^{m} \lambda_i^* M(\bar{\varphi}_{z_i,\alpha_i}, |\Delta| f_i), \tag{13}$$

with $\lambda_i^* := \max\{\lambda_i : \lambda \in \Lambda\}$ and $z := Tx$. It turns out that for periodically monotone functions φ (including $\bar{\varphi}_{z_i,\alpha_i}$ and $\hat{\varphi}_{z_i}$) exact expressions of $M(\varphi, B)$ can be obtained; in all other cases an upper bound is available. Moreover, as shown in Examples 1 and 2 of Romeijnders et al. (2016) we have for every $z_i, \alpha_i \in \mathbb{R}$ and for every $B \in \mathbb{R}$ with $B > 0$ that

$$M(\bar{\varphi}_{z_i,\alpha_i}, B) \leq h(B) \qquad \text{and} \qquad M(\hat{\varphi}_{z_i}, B) = M(-\hat{\varphi}_{z_i}, B) = \frac{1}{2} h(B) \tag{14}$$

with h defined in (7). Combining (13) and the first inequality in (14) we obtain the error bound from Theorem 1. Moreover, observe that the difference of a factor 2 between $M(\bar{\varphi}_{z_i,\alpha_i}, B)$ and $M(\hat{\varphi}_{z_i}, B)$ in (14) causes the factor 2 difference between the error bounds of the α-approximations and the shifted LP-relaxation approximation.

Since the error bounds in Theorem 1 are determined using worst-case analysis, among others in the form of (12), the question arises how sharp these error bounds actually are. As already mentioned in the introduction they are reasonably tight when compared with $\|G - G_\alpha\|_\infty$ and $\|G - \hat{G}\|_\infty$. In Sections 3–5, however, we will compare the error bounds with $G(x_\alpha) - \eta^*$ and $G(\hat{x}) - \eta^*$ and show that the quality of the convex approximations may in fact be much better than Theorem 1 guarantees. That is, in some important cases the worst-case error bounds are not very sharp.

2.2. Assessing the quality of candidate solutions using sampling

In this subsection we review sampling methods for assessing the quality of candidate solutions, $x \in X$, for model (1). In particular, we discuss the multiple replications procedure (MRP) of Mak et al. (1999). This procedure is easy to implement and works under very general assumptions, which are satisfied by the integer recourse function g defined in (2), at least if $g(x, \omega)$ has finite variance for all $x \in X$. In contrast, since $g(\cdot, \omega)$ is generally not continuous for every $\omega \in \Omega$ if the second-stage involves integer decision variables, we cannot use the single- or two-replications procedures in Bayraksan and Morton (2006).

The MRP can be applied to any candidate solution, $x \in X$, independent of the method by which x is obtained. So, for the TU integer recourse models that we consider in Sections 3–5 we can use the MRP with $x := x_\alpha$ and $x := \hat{x}$, where x_α and $\hat{x}$ denote solutions of the α-approximations and shifted LP-relaxation approximation, respectively.

Other sampling methods for assessing solution quality in stochastic programming problems include Glynn and Infanger (2013), Higle and Sen (1991, 1996), and Shapiro and Homem-de-Mello (1998); see also the tutorial of Bayraksan and Morton (2009). Although descriptions of MRP can be found in this tutorial and in, e.g., Bayraksan and Morton (2006), we discuss it here to set notation for what follows.

2.2.1. Multiple replications procedure. We measure the quality of a candidate solution, $x \in X$, of (1) by its optimality gap $\theta(x) := G(x) - \eta^*$. This gap cannot be obtained by straightforward computation, since it is typically impossible to calculate η^* exactly and because evaluating $G(x)$ can require computing a higher-dimensional integral. Nonetheless, the optimality gap may be estimated using sampling.

Let $\omega^1, \ldots, \omega^n$ denote an i.i.d. sample from the distribution of ω. Then, $n^{-1} \sum_{j=1}^n g(x, \omega^j)$ is a consistent estimator of $G(x)$. We can estimate η^* by solving

$$(SP_n) \quad \eta_n^* := \min_x \left\{ \frac{1}{n} \sum_{j=1}^n g(x, \omega^j) : x \in X \right\}. \tag{15}$$

Model (15) is of the same form as model (1), but will be computationally tractable if the sample size is small enough. The estimator η_n^* of η^* has a negative bias, i.e., $\mathbb{E}[\eta_n^*] \leq \eta^*$; see Mak et al. (1999). In this way, $\theta_n(x) := n^{-1} \sum_{j=1}^n g(x, \omega^j) - \eta_n^*$ will be a conservative estimate of the optimality gap $\theta(x)$ in the sense that $\mathbb{E}[\theta_n(x)] \geq \theta(x)$.

The distribution of $\theta_n(x)$ may be asymptotically non-normal, making it more difficult to derive probabilistic statements on $\theta_n(x)$. This issue is circumvented by replicating the procedure N_r times and applying the central limit theorem (CLT). A complete description of the MRP is given below.

Multiple Replications Procedure:

Step 1: For $i = 1, \ldots, N_r$,

(i) Sample (i.i.d.) observations $\omega^{i1}, \ldots, \omega^{in}$ from the distribution of ω.

(ii) Solve (SP_n) in (15) using $\omega^{i1}, \ldots, \omega^{in}$ yielding objective η_n^{i*} and solution x_n^{i*}.

(iii) Calculate $\theta_n^i(x) := n^{-1} \sum_{j=1}^n (g(x, \omega^{ij}) - g(x_n^{i*}, \omega^{ij}))$.

Step 2: Calculate the gap estimate $\bar{\theta}_n(x)$ and sample variance $s_\theta^2(x)$ by

$$\bar{\theta}_n(x) := \frac{1}{N_r} \sum_{i=1}^{N_r} \theta_n^i(x) \quad \text{and} \quad s_\theta^2(x) = \frac{1}{N_r - 1} \sum_{i=1}^{N_r} (\theta_n^i(x) - \bar{\theta}_n(x))^2.$$

Step 3: Let $\epsilon_\theta := t_{N_r-1,\gamma} \cdot s_\theta(x)/\sqrt{N_r}$, where $t_{N_r-1,\gamma}$ denotes the $(1-\gamma)\%$ quantile of the t distribution with $N_r - 1$ degrees of freedom. Then, the one-sided $(1-\gamma)\%$ confidence interval on $\theta(x) = G(x) - \eta^*$ is given by $[0, \bar{\theta}_n(x) + \epsilon_\theta]$. That is, if the CLT were to hold exactly for finite sample size, N_r, we would have

$$\mathbb{P}\Big\{G(x) - \eta^* \in [0, \bar{\theta}_n(x) + \epsilon_\theta]\Big\} \geq 1 - \gamma.$$

Step 1(i) of the MRP need not use an i.i.d. sample; only i.i.d. samples over replications i are required. For example, throughout this paper we use Latin hypercube sampling (LHS) in this step, which reduces variance and also often decreases the bias. This is important because the width of the confidence interval of $\theta(x)$ may be large since (i) x is suboptimal, (ii) the negative bias of $\bar{\theta}_n(x)$ is large, or (iii) the sample variance, and thus ϵ_θ, is large. Using LHS we reduce the effect of (ii) and (iii) so that we can better assess the quality of the candidate solution x; see, e.g., Freimer et al. (2012).

Although the purpose of the MRP is to assess the quality of a candidate solution, $x \in X$, it also calculates potential candidate solutions x_n^{i*} in Step 1(ii), and can thus also be considered a sampling (solution) method. The candidate solutions will most likely be suboptimal, particularly when the sample size n is small, but we can obtain the best among them using an out-of-sample evaluation or by averaging them (Sen and Liu 2014) if X is convex. In Sections 3–5 we compare the solution of this sampling method with the solutions obtained from the convex approximations.

2.2.2. Performance measures. In Sections 3–5, we report three performance measures, ρ_1, ρ_2, and ρ_3, which respectively correspond to our three main contributions, enumerated in Section 1. Again, we use $\hat{x}$ to denote an optimal solution to the convex approximation of model (3).

The first performance measure,

$$\rho_1(\hat{x}) := \frac{\bar{\theta}_n(\hat{x}) + \epsilon_\theta}{2U(G, \hat{G})} \times 100\%,$$

compares the optimality gap, or more precisely the width of the MRP's $(1-\gamma)$-level confidence interval on the optimality gap, $G(\hat{x}) - \eta^*$, with the error bound $2U(G, \hat{G})$ of Theorem 1 and inequality (4). Thus, $\rho_1(\hat{x})$ estimates the sharpness of the error bound. If $\rho_1(\hat{x})$ is small, then the actual performance of the convex approximation is better than its error bound suggests.

The second performance measure,

$$\rho_2(\hat{x}) := \frac{\bar{\theta}_n(\hat{x}) + \epsilon_\theta}{N_r^{-1}\sum_{i=1}^{N_r} \eta_n^{i*}} \times 100\%, \tag{16}$$

compares the same estimate of the optimality gap with an estimate of the optimal objective value, η^*, provided by the MRP. Thus, $\rho_2(\hat{x})$ measures the quality of the approximating solution $\hat{x}$, which is estimated to be a good solution if $\rho_2(\hat{x})$ is small. As discussed in Section 2.2.1, $\mathbb{E}[\eta_n^{i*}] \leq \eta^*$ and $\mathbb{E}[\bar{\theta}_n(\hat{x}) + \epsilon_\theta] \geq G(\hat{x}) - \eta^*$, so that the performance measure $\rho_2(\hat{x})$ is a conservative estimate of $(G(\hat{x}) - \eta^*)/\eta^* \times 100\%$.

Furthermore, we compare the approximating solution $\hat{x}$ with a sampling solution x^S. This sampling solution is the best of the solutions x_n^{i*}, $i = 1, \ldots, N_r$, obtained during the MRP, and their average $N_r^{-1}\sum_{i=1}^{N_r} x_n^{i*}$, when X is convex. To assess the quality of the sampling solution we report $\rho_2(x^S)$, and to compare $\hat{x}$ and x^S we use the performance measure

$$\rho_3(\hat{x}, x^S) := \frac{G(\hat{x}) - G(x^S)}{N_r^{-1}\sum_{i=1}^{N_r} \eta_n^{i*}} \times 100\%.$$

We note that we essentially use the MRP twice; first just to obtain the sampling solution, x^S, and second to assess the quality of $\hat{x}$ and x^S.

3. Integer newsvendor problem

3.1. Problem definition and analysis

In this section we consider an integer newsvendor problem. This problem, which is an example of a model with simple integer recourse, is the simplest version of a TU integer recourse problem with g as in (2), and is defined as

$$\eta^* = \min_{x \geq 0} \left\{ cx + r\mathbb{E}_\omega[\lceil \omega - x \rceil^+] \right\}, \tag{17}$$

with $0 < c < r$ and $\lceil s \rceil^+ := \max\{0, \lceil s \rceil\}$, $s \in \mathbb{R}$. We have substituted the exact expression

$$g(x, \omega) = cx + \min_y \{ry : y \geq \omega - x,\, y \in \mathbb{Z}_+\} = cx + r\lceil \omega - x \rceil^+, \quad x \geq 0, \omega \in \mathbb{R},$$

in the objective function of (17). Moreover, observe that the problem is generally non-convex because of the round-up operator.

The approximating models corresponding to the α-approximations and shifted LP-relaxation approximation defined in (5) and (6), respectively, reduce to

$$\eta_\alpha := \min_{x \geq 0} \left\{ cx + r\mathbb{E}_\omega[(\lceil \omega \rceil_\alpha - x)^+] \right\}, \tag{18}$$

and

$$\hat{\eta} := \min_{x \geq 0} \left\{ cx + r\mathbb{E}_\omega[(\omega + 1/2 - x)^+] \right\}. \tag{19}$$

Even though the integer newsvendor problem is simple, we consider it here because it allows for a more precise analysis than we can perform in Sections 4 and 5. For the models in (18) and (19) we can obtain closed-form solutions. For example, for the shifted LP-relaxation approximation we have $\hat{x} = (1/2 + F^{-1}(\frac{r-c}{r}))^+$ with F^{-1} denoting the quantile function of ω. The quality of these solutions is guaranteed by the error bounds in Theorem 1. Combining those with (4) we have for this integer newsvendor problem that

$$G(x_\alpha) - \eta^* \leq 2rh(|\Delta|f) \quad \text{and} \quad G(\hat{x}) - \eta^* \leq rh(|\Delta|f),$$

where h is defined in (7). In the next subsection we analyze the sharpness of these bounds using numerical experiments. Below we compute these bounds for both normal and log-normal random variables ω.

Example 1. Let $\omega \sim N(\mu, \sigma^2)$ be a normally distributed random variable with pdf f. Then, f is unimodal with maximum $1/\sqrt{2\pi\sigma^2}$ at $x = \mu$, so that $|\Delta|f = 2/\sqrt{2\pi\sigma^2} = \sigma^{-1}\sqrt{2/\pi}$, and thus

$$G(x_\alpha) - \eta^* \leq 2rh(\sigma^{-1}\sqrt{2/\pi}) \quad \text{and} \quad G(\hat{x}) - \eta^* \leq rh(\sigma^{-1}\sqrt{2/\pi}),$$

where

$$h(\sigma^{-1}\sqrt{2/\pi}) = \begin{cases} 1 - \sigma\sqrt{2\pi}, & \sigma \leq 1/\sqrt{8\pi}, \\ (8\sigma)^{-1}\sqrt{2/\pi}, & \sigma \geq 1/\sqrt{8\pi}. \end{cases}$$

Similarly, we have $\|G - G_\alpha\|_\infty \leq rh(\sigma^{-1}\sqrt{2/\pi})$ and $\|G - \hat{G}\|_\infty \leq \frac{1}{2}rh(\sigma^{-1}\sqrt{2/\pi})$. Figure 1 compares the actual values of the supremum norms with their upper bound for the case $r = 1$. It is the same figure as in Example 5.11 of Romeijnders et al. (2015), but now with the values of $\|G - \hat{G}\|_\infty$ included. Observe that indeed the upper bound is reasonably tight, and that the shifted LP-relaxation approximation is better than the α-approximations.

Example 2. Let ω be lognormally distributed, i.e., $\ln\omega \sim N(\mu, \sigma^2)$. In this case, $\mathbb{E}_\omega[\omega] = \exp\{\mu + \sigma^2/2\}$ and $\mathrm{Var}(\omega) = (\exp\{\sigma^2\} - 1)\exp\{2\mu + \sigma^2\}$, and moreover ω has pdf

$$f(x) = \begin{cases} \dfrac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{ -\dfrac{1}{2}\left(\dfrac{\ln x - \mu}{\sigma}\right)^2 \right\}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

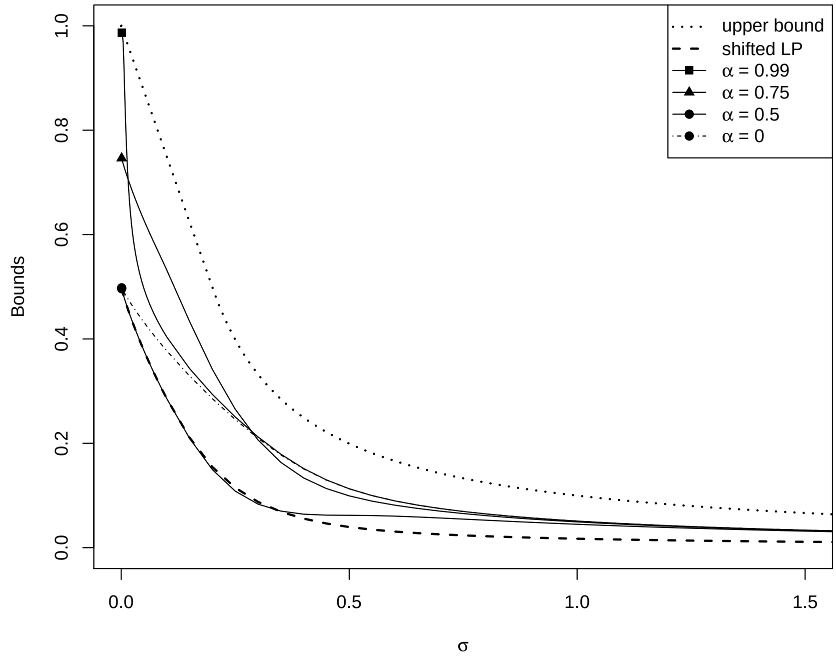

Figure 1 **The supremum norms $\|G-\hat{G}\|_\infty$ and $\|G-G_\alpha\|_\infty$, and their upper bound $rh(|\Delta|f)$, of Example 1 (with $r=1$) as a function of σ, the standard deviation of the random variable $\omega\sim N(0,\sigma^2)$. The dotted line corresponds to $h(|\Delta|f)$, the dashed line to $\|G-\hat{G}\|_\infty$, and the remaining four lines to $\|G-G_\alpha\|_\infty$ for $\alpha=0,0.5,0.75,0.99$.**

The pdf f is unimodal with mode $\exp\{\mu-\sigma^2\}$. It follows immediately that

$$|\Delta|f=\sqrt{\frac{2}{\pi\sigma^2}}\exp\left\{\frac{1}{2}\sigma^2-\mu\right\}.$$

In contrast to the normal case, the total variation $|\Delta|f$ of f depends on μ, and it decreases as μ increases. Moreover, for large values of σ the total variation $|\Delta|f$ is also large in the lognormal case, even though the variance of ω is large. This illustrates that there is not necessarily a one-to-one relation between the variance of ω and the total variation $|\Delta|f$ of the pdf, f, of ω, as is the case when ω is normally distributed.

Remark 2. In general it is hard to compute $\|G-G_\alpha\|_\infty$ and $\|G-\hat{G}\|_\infty$, but for this integer newsvendor problem it is possible using brute force computation. In fact, values of $G(x_\alpha)-\eta^*$ and $G(\hat{x})-\eta^*$ might also be obtained in a similar way. However, we prefer to use the MRP instead for better comparison with the problems of Sections 4 and 5 for which brute force computations are intractable.

3.2. Numerical experiments

Here, we carry out numerical experiments for the integer newsvendor problem. We compare the performance of x_α and $\hat{x}$, the approximating solutions of the α-approximation and shifted LP-relaxation approximation, respectively. We also apply the MRP to estimate the optimality gaps $\theta(x_\alpha) = G(x_\alpha) - \eta^*$ and $\theta(\hat{x}) = G(\hat{x}) - \eta^*$, and we compare these optimality gaps, or rather their estimates, with the upper bounds $2rh(|\Delta|f)$ and $rh(|\Delta|f)$.

We consider two types of distributions, the normal ($\omega \sim N(\mu, \sigma^2)$) and lognormal ($\ln \omega \sim N(\mu, \sigma^2)$). The latter is a distribution with a heavy tail, whereas the tails of the normal distribution decrease exponentially. The value of c is standardized to 1 and we use $r \in \{1.05, 1.3, 2, 4, 20\}$ in our experiments. The values of r are chosen such that (approximately) $\frac{r-c}{r} \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$, and thus the solution $\hat{x}$ is obtained by computing very different quantiles of the distribution of ω.

Table 1 Comparison of the shifted LP-relaxation approximation and the α-approximation for the integer newsvendor problem (17) when ω is normally distributed. Exact objective values $G(\hat{x})$ and $G(x_\alpha)$ with $\alpha = 0, 0.25, 0.5$, and 0.75 are given; for each experiment the minimum of these objective values is displayed in bold.

Exp.	μ	σ	r	$G(\hat{x})$	$G(x_0)$	$G(x_{0.25})$	$G(x_{0.5})$	$G(x_{0.75})$
1	1	0.1	1.05	1.336	1.526	**1.257**	1.500	1.750
2	1	0.1	1.3	1.433	1.667	**1.258**	1.500	1.750
3	1	0.1	2	1.500	2.000	**1.262**	1.500	1.750
4	1	0.1	4	1.567	2.000	**1.275**	1.500	1.750
5	1	0.1	20	1.664	2.000	**1.374**	1.500	1.750
6	1	0.5	1.05	1.550	1.550	1.564	1.554	**1.548**
7	1	0.5	1.3	1.673	1.697	**1.670**	1.713	1.761
8	1	0.5	2	**1.820**	2.046	1.880	**1.820**	1.884
9	1	0.5	4	2.026	2.091	2.275	2.140	**2.018**
10	1	0.5	20	2.404	2.456	**2.374**	2.527	2.755
11	1	1	1.05	**1.604**	**1.604**	1.611	**1.604**	**1.604**
12	1	1	1.3	**1.906**	1.910	1.943	1.931	1.908
13	1	1	2	**2.264**	2.366	2.290	**2.264**	2.290
14	1	1	4	**2.717**	2.731	2.724	2.793	2.829
15	1	1	20	**3.481**	3.482	3.506	3.629	3.613
16	1	3	1.05	**2.198**	**2.198**	**2.198**	**2.198**	**2.198**
17	1	3	1.3	**2.785**	**2.785**	**2.785**	**2.785**	**2.785**
18	1	3	2	**3.883**	3.916	3.891	**3.883**	3.891
19	1	3	4	**5.296**	5.344	5.311	**5.296**	5.307
20	1	3	20	**7.660**	7.723	7.669	7.662	7.697
21	1	10	1.05	**5.034**	**5.034**	**5.034**	**5.034**	**5.034**
22	1	10	1.3	**6.377**	**6.377**	**6.377**	**6.377**	**6.377**
23	1	10	2	**9.476**	9.485	9.478	**9.476**	9.478
24	1	10	4	**14.206**	14.210	**14.206**	14.210	14.221
25	1	10	20	**22.119**	**22.119**	22.128	22.139	22.122

Table 1 compares the shifted LP-relaxation approximation and the α-approximation with $\alpha = 0, 0.25, 0.5$, and 0.75 for ω normally distributed. For large values of σ, i.e., $\sigma = 10$, the difference between the approximations is very small, whereas for small values of σ, i.e., $\sigma = 0.1$, the approximations differ significantly. In the latter case, the solution $x_{0.25}$ is best. In some sense, this can be considered a coincidence because by construction the optimal solution to (18) is either $x_\alpha = 0$ or $x_\alpha \in \alpha + \mathbb{Z}$, and $x_{0.25} = 1.25$ is close to the optimal solution of the integer newsvendor problem (17). On the other hand, for medium to large values of σ, the shifted LP-relaxation approximation outperforms all α-approximations, in line with the fact that its error bound is better by a factor of 2. We thus prefer the shifted LP-relaxation approximation, also because in contrast to the α-approximation it does not require specification of parameter α. For the α-approximations the experiments suggest that it is important to select a good value of α, in particular if σ is small, but this value of α depends on the fractional value of the unknown optimal solution x^* of (17). The analog of Table 1 when ω is lognormal is similar, and so we do not include those results here.

Table 1 does not give any information on how close to optimal the approximations are. So, we use the MRP to evaluate the optimality gaps. We restrict our attention here to the shifted LP-relaxation approximation. Results for the α-approximations are very similar.

Table 2 Numerical results for the shifted LP-relaxation approximation applied to the integer newsvendor problem (17) with ω normally distributed. The MRP is applied with $N_r = 30$, $n = 1000$, and $\gamma = 0.05$.

			Optimality gap compared with upper bound (in %): $\rho_1(\hat{x})$					Optimality gap compared with optimal objective (in %): $\rho_2(\hat{x})$				
			r					r				
μ	σ	$\|\Delta\|f$	1.05	1.3	2	4	20	1.05	1.3	2	4	20
1	0.1	7.9	15.2	20.6	17.1	10.0	2.3	9.9	16.8	20.6	23.5	26.8
1	0.5	1.5	1.9	2.0	0.6	1.6	1.5	0.3	0.3	0.1	0.6	2.5
1	1	0.8	2.5	2.2	1.9	1.7	1.3	0.2	0.2	0.2	0.2	0.7
1	3	0.26	2.2	5.6	10.6	8.0	5.8	0.03	0.1	0.2	0.2	0.5
1	10	0.08	9.4	11.1	61.0	46.3	32.7	0.02	0.02	0.1	0.1	0.3

For the MRP we use $N_r = 30$, $n = 1000$, and $\gamma = 0.05$, and we use LHS in Step 1(i) of the procedure. The approximating solutions $\hat{x}$ are obtained by solving (19) exactly. Alternatively, this solution could also have been obtained using a sample average approximation of (19) with a large sample. We report three performance measures ρ_1, ρ_2, and ρ_3, discussed in Section 2.2.2.

In Table 2 we show the performance measures $\rho_1(\hat{x})$ and $\rho_2(\hat{x})$ for normally distributed ω and in Table 3 for lognormal ω. For the normal case, we observe that the approximating solution $\hat{x}$ is good in case of medium and high variability (i.e., $\sigma = 0.5, 1, 3$, and $\sigma = 10$,

Table 3 Numerical results for the shifted LP-relaxation approximation applied to the integer newsvendor problem (17) with ω lognormally distributed. The MRP is applied with $N_r = 30$, $n = 1000$, and $\gamma = 0.05$.

			Optimality gap compared with upper bound (in %): $\rho_1(\hat{x})$					Optimality gap compared with optimal objective (in %): $\rho_2(\hat{x})$				
			r					r				
μ	σ	$\lvert\Delta\rvert f$	1.05	1.3	2	4	20	1.05	1.3	2	4	20
0	0.1	8.0	15.2	19.1	15.7	9.1	2.1	9.7	15.4	18.6	20.9	23.6
0	0.5	1.8	2.8	0.9	1.3	1.3	0.7	0.4	0.1	0.3	0.5	0.9
0	1.5	1.6	2.6	2.6	2.3	1.9	2.0	0.2	0.2	0.2	0.2	0.3
1	0.5	0.67	5.0	3.8	3.1	3.0	2.5	0.1	0.1	0.1	0.2	0.5
2	1.7	0.26	30.9	33.5	31.8	32.4	15.1	0.03	0.04	0.04	0.04	0.03

respectively). Indeed, $\rho_2(\hat{x})$ suggests that, with high confidence, the objective value of the solution $\hat{x}$ is within 1% of the optimal objective function value in almost all cases. In contrast, for low variability ($\sigma = 0.1$) the solution $\hat{x}$ is not good: $\rho_2(\hat{x})$ may exceed 20%. This is in line with the error bound of Theorem 1, which is larger if σ is smaller. The values of $\rho_1(\hat{x})$, however, are only small in case of medium variability. In these cases, the error bound of Theorem 1 is not sharp: the quality of the solution $\hat{x}$ is much better than the error bound suggests. For low and high variability the error bound is reasonably sharp: the value of $\rho_1(\hat{x})$ may be above 15% and may even range up to 60%. In case of high variability, these large values of $\rho_1(\hat{x})$ are inherent to the nature of the total variation error bound and the MRP optimality gap. Indeed, as the standard deviation σ grows, the total variation error bound shrinks, whereas the MRP optimality gap remains approximately the same.

For the lognormal case, we obtain similar results (see Table 3). We have selected values of μ and σ so that $|\Delta|f$ approximately matches those in the normal case. As detailed in Example 2 this does not mean that the variances of ω match, however. For example, in the lognormal case with $\mu = 0$ and $\sigma = 1.5$, we have $\mathrm{Var}(\omega) \approx 80.5$, and for $\mu = 1$ and $\sigma = 0.5$, we have $\mathrm{Var}(\omega) \approx 2.7$.

Comparing the shifted LP-relaxation approximation solution, $\hat{x}$, and the sampling solution, x^S, in Table 4 for normal random variables, ω, we observe that $\rho_3(\hat{x}, x^S)$ is only large in the case of low variability (i.e., $\sigma = 0.1$). Indeed, in contrast to the shifted LP-relaxation approximation, the sampling solution, x^S, is good in the low variability case as well. In fact, $\rho_2(x^S)$ is small in all cases. This is as expected, since we use a large sample (of size $n = 1000$) to obtain the sampling solution, x^S. In some higher-dimensional problems, larger sample sizes are required to obtain high-quality solutions, and yet such problems are more difficult to solve and could be intractable even with modest sample sizes. With this in mind, the shifted LP-relaxation approximation performs well in case of medium to high variability. In these cases both solution methods perform approximately the same. Since we obtain similar results for the lognormal case, we omit those computational results here.

Table 4 A comparison between the shifted LP-relaxation approximation and a sampling solution method for the integer newsvendor problem (17) with ω normally distributed. The MRP is applied with $N_r = 30$, $n = 1000$, and $\gamma = 0.05$. Values reported as $\pm$ 0.00 are small in magnitude while values reported as 0 are indeed zero.

			Difference between sampling and approximation (in %): $\rho_3(\hat{x}, x^S)$					Optimality gap compared with optimal objective (in %): $\rho_2(x^S)$				
			r					r				
μ	σ	$\|\Delta\|f$	1.05	1.3	2	4	20	1.05	1.3	2	4	20
1	0.1	7.9	9.8	16.7	20.5	23.2	25.7	0.08	0.1	0.2	0.3	1.1
1	0.5	1.5	0.15	0.18	0.00	0.4	1.8	0.1	0.1	0.1	0.2	0.8
1	1	0.8	-0.00	0.01	-0.00	0.02	0.15	0.2	0.2	0.2	0.2	0.7
1	3	0.26	0	0	-0.00	0.00	0.002	0.04	0.1	0.2	0.2	0.5
1	10	0.08	0	0	-0.00	-0.00	-0.00	0.02	0.02	0.1	0.1	0.3

4. Fleet allocation and routing problem

This section discusses a variant of the fleet allocation and routing problem introduced in Donohue and Birge (1995). Mak et al. (1999) also report numerical results for this problem.

The problem may be viewed as a two-stage totally unimodular integer recourse model, but with relatively complete recourse instead of complete recourse and with deterministic side constraints in the second stage. This is why we have to reconsider what type of convex approximations are suitable for this problem, and, moreover, why we (again) have to derive an error bound for these approximations.

First, we define the problem, formulate it as a two-stage integer recourse model, and derive properties of this model in Section 4.1. Next, in Section 4.2 we construct a concave approximation g_0 —since we are maximizing— of the recourse function g, and we derive an error bound for this approximation. Finally, in Section 4.3 we carry out numerical experiments comparing the actual error of the concave approximation with its error bound.

4.1. Problem definition and model formulation

Section 4.1.1 defines the problem. Then, Section 4.1.2 formulates the stochastic integer program, and Section 4.1.3 discusses properties of the model.

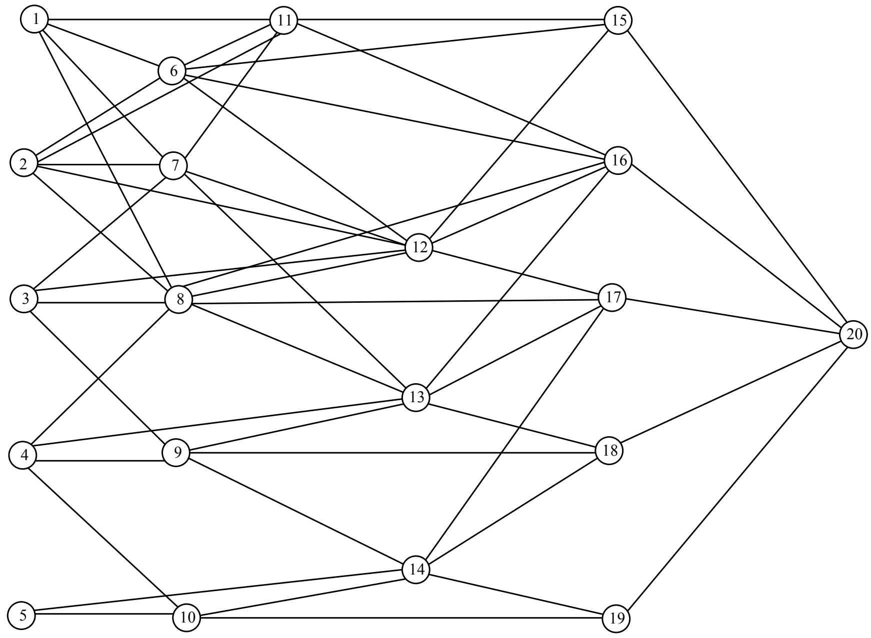

Figure 2 **The graph $\mathcal{G} = (V, E)$ used for the numerical experiments. Nodes 1-5 are source nodes (V_s) and node 20 is the sink node (t). All arcs are directed from left to right.**

4.1.1. Problem definition. Consider an acyclic directed graph $\mathcal{G} = (V, E)$, modeling a road network. A fleet of trucks will traverse this network starting at source nodes $V_s \subset V$ and finishing at a sink node $t \in V$. For every arc $(i, j) \in E$, the first D_{ij} trucks traversing the arc receive a reward $r_{ij} > 0$ and subsequent trucks incur a cost $c_{ij} > 0$. The quantities D_{ij} are "soft" demands, or customers requesting service, along arc (i, j). Trucks receive profit if they serve a customer, and incur a cost otherwise. The problem is to allocate N trucks to the source nodes V_s and route them through the network to maximize profit.

When we allocate trucks to the source nodes, the demands D_{ij} are unknown. We assume that they are in part random but may be increased by investments, or marketing actions β_{ij}, which incur unit costs q_{ij}. That is, the demand $D_{ij}(\omega_{ij}, \beta_{ij})$ is a function of the random variable ω_{ij} (with known probability distribution) and the investments β_{ij}. The effect of the investments may be additive ($D_{ij}(\omega_{ij}, \beta_{ij}) = \omega_{ij} + \beta_{ij}$) or multiplicative ($D_{ij}(\omega_{ij}, \beta_{ij}) = \omega_{ij}(1 + \beta_{ij})$). In either case, when the investment is zero, we have $D_{ij}(\omega_{ij}, 0) = \omega_{ij}$. Observe that we do not define demands to be integer; instead we will impose integrality restrictions on the number of trucks traversing an arc. Thus, our objective is to allocate the trucks to the source nodes *and* invest β in the arcs at costs q to maximize *expected* profits.

4.1.2. Model formulation: two-stage recourse. This problem can be formulated as a two-stage integer recourse model. In the first stage we decide the number of trucks, n_i, to allocate to each source node, $i \in V_s$, and we decide the investments, $\beta_{ij} \geq 0$, for each arc, $(i,j) \in E$. In the second stage we let $y_{ij} \in \mathbb{Z}_+$ denote the number of trucks receiving a reward r_{ij} when traversing arc (i,j), and we let $z_{ij} \in \mathbb{Z}_+$ denote the number of trucks incurring a cost c_{ij}. Then, the two-stage integer recourse model for this problem is given by

$$\eta^* := \max_x \Big\{ \mathbb{E}_\omega[g(x,\omega)] : x \in X \Big\}, \tag{20}$$

where $x := (n,\beta)$ with feasible region $X := \{(n,\beta) \in \mathbb{Z}_+^{|V_s|} \times \mathbb{R}_+^{|E|} : \sum_{i \in V_s} n_i = N\}$, and

$$\begin{aligned} g(x,\omega) := -q\beta + \max_{y,z} \quad & ry - cz \\ \text{s.t.} \quad & \sum_{j:(i,j)\in E} (y_{ij} + z_{ij}) = n_i & i \in V_s \qquad (21) \\ & \sum_{j:(i,j)\in E} (y_{ij} + z_{ij}) - \sum_{j:(j,i)\in E} (y_{ji} + z_{ji}) = 0 & i \in V \backslash (V_s \cup \{t\}) \qquad (22) \\ & \sum_{j:(j,i)\in E} (y_{ji} + z_{ji}) = N & i = t \qquad (23) \\ & 0 \leq y_{ij} \leq D_{ij}(\omega_{ij}, \beta_{ij}), \quad 0 \leq z_{ij} & (i,j) \in E \\ & y, z \in \mathbb{Z}^{|E|}. \end{aligned}$$

Here, constraints (21)–(23) represent flow balance constraints, modeling, respectively, that n_i trucks must leave source node $i \in V_s$, that every truck that enters node $i \in V \backslash (V_s \cup \{t\})$ must leave that node, and that all N trucks must arrive at sink node t.

Remark 3. This fleet allocation and routing model is a special case of the two-stage integer recourse model defined in (2), since maximization can easily be reformulated as minimization and the flow balance constraints can be captured by the set Y. When the effect of investments is only additive, there is only randomness in the right-hand side vector $\zeta(\omega)$. If multiplicative effects are considered, then also the technology matrix $T(\omega)$ is random.

Let A denote the node-arc incidence matrix of $\mathcal{G}$, where the rows of A correspond to the nodes of $\mathcal{G}$ and the columns of A to the arcs of $\mathcal{G}$. The column of A corresponding to arc $(i,j) \in E$ has one entry equal to $+1$ in row i, one entry equal to -1 in row j, and the remaining entries equal zero. Defining $b(n) = (n, 0, -N)$, the flow balance constraints can be written as $Ay + Az = b(n)$. Thus, for every $x = (n,\beta) \in X$,

$$g(x,\omega) = -q\beta + \max_{y,z} \Big\{ ry - cz : Ay + Az = b(n),\, y \leq D(\omega,\beta),\, y,z \in \mathbb{Z}_+^{|E|} \Big\}. \tag{24}$$

4.1.3. Properties of the recourse function g. Here, we discuss properties of the recourse function g. We assume that the directed acyclic graph $\mathcal{G}$ is *t-connected*; i.e., we assume that for every node i in the graph there is a directed i-t path. Under this assumption, we show that the recourse is relatively complete and that the recourse matrix corresponding to the second-stage optimization problem in g is TU.

LEMMA 1. *Let graph $\mathcal{G} = (V, E)$ be t-connected, let g be the recourse function defined in (24), and let $G(x) := \mathbb{E}_\omega[g(x,\omega)]$, $x \in X$. Then, the recourse is* relatively complete *and* sufficiently expensive*; that is,*

(i) $g(x,\omega)$ is finite for every $x \in X$ and $\omega \in \mathbb{R}_+^{|E|}$; and,

(ii) $G(x)$ is finite for every $x \in X$ and nonnegative random vector ω.

Proof: Let $x \in X$ and $\omega \in \mathbb{R}_+^{|E|}$ be given. Clearly, there exists a feasible solution of the maximization problem in g, for example using $y = 0$ and z such that (y,z) is feasible. Moreover, since the graph $\mathcal{G}$ is acyclic, it follows immediately from the flow balance constraints that for any feasible solution we have $0 \leq y_{ij}, z_{ij} \leq N$ for all $(i,j) \in E$, and thus

$$-N \sum_{(i,j)\in E} c_{ij} \leq g(x,\omega) \leq N \sum_{(i,j)\in E} r_{ij},$$

implying that (i) $g(x,\omega)$ is finite. Now it is not hard to prove (ii) since for every $x \in X$ and nonnegative random vector ω we have

$$-N \sum_{(i,j)\in E} c_{ij} \leq G(x) \leq N \sum_{(i,j)\in E} r_{ij}. \quad \square$$

REMARK 4. The recourse matrix W of the maximization problem in g, defined as

$$W = \begin{bmatrix} A & A \\ I & 0 \end{bmatrix},$$

is TU (e.g., Schrijver 1986), implying that we can represent the maximization problem as a linear program. This does not imply, however, that (20) is as easy to solve as a two-stage continuous recourse problem since the LP representation of g involves rounding down the demands $D_{ij}(\omega_{ij}, \beta_{ij})$, similar to (8) in minimization problems.

In what follows we use that W is TU to obtain a dual representation of g.

4.2. Concave approximation

Here, we derive a concave approximation of g (since we are maximizing instead of minimizing) using the same type of approximations as in Section 2.1.1, i.e., an α-approximation and a shifted LP-relaxation approximation. However, difficulties arise because the recourse is relatively complete rather than complete. Moreover, to derive an error bound we have to extend the analysis in Section 2.1.3 to be able to deal with the flow balance constraints, which we may view as deterministic second-stage side constraints. We will do so by using the same line of proof as in Section 2.1.3, exploiting the results presented there.

4.2.1. Definition of the concave approximation g_0. The main idea in the convex approximations of Section 2.1.1 is to simultaneously relax the integrality constraints and replace the right-hand side ω by $\lceil \omega \rceil_\alpha$ or $\omega + 1/2e_m$. In this case, since we are maximizing instead of minimizing, and the right-hand side $D(\omega, \beta)$ is rounded down instead of up, we analogously replace ω by either $\lfloor \omega \rfloor_\alpha := \lfloor \omega - \alpha \rfloor + \alpha$ or $\omega - 1/2e_m$.

However, $\omega - 1/2e_m$ may be negative with positive probability, even if the random vector $\omega \geq 0$, which implies that for $\beta = 0$, or small, the approximating demands may be negative, and thus the approximating maximization problem infeasible. The same holds for $\lfloor \omega \rfloor_\alpha$, unless $\alpha \in \mathbb{Z}^m$. For such $\alpha \in \mathbb{Z}^m$, we have that $\lfloor \omega \rfloor_\alpha = \lfloor \omega \rfloor \geq 0$ if $\omega \geq 0$.

Thus, interestingly, the only reasonable concave approximation that can be used for every nonnegative random vector ω is an α-approximation with $\alpha = 0$. This approximation, denoted g_0, is defined for every $x \in X$ and $\omega \in \mathbb{R}_+^{|E|}$ as

$$g_0(x, \omega) = -q\beta + \max_{y,z} \left\{ ry - cz : Ay + Az = b(n),\ y \leq \tilde{D}(\omega, \beta),\ y, z \in \mathbb{R}_+^{|E|} \right\}, \tag{25}$$

where for every $(i, j) \in E$,

$$\tilde{D}_{ij}(\omega_{ij}, \beta_{ij}) := \begin{cases} \lfloor \omega_{ij} \rfloor + \omega_{ij}\beta_{ij}, & (i,j) \in E^*, \\ \lfloor \omega_{ij} \rfloor + \beta_{ij}, & (i,j) \in E^+, \\ \lfloor \omega_{ij} \rfloor, & (i,j) \in E^0, \end{cases}$$

and where E^*, E^+, and E^0 partition E into subsets with multiplicative, additive, and no investment effects, respectively. Observe that $g_0(x, \omega)$ is concave in x for every $\omega \in \mathbb{R}_+^{|E|}$. Moreover, notice that for multiplicative investments effects we have $\tilde{D}_{ij}(\omega_{ij}, \beta_{ij}) \neq D_{ij}(\lfloor \omega_{ij} \rfloor, \beta_{ij})$ unless $\omega_{ij} \in \mathbb{Z}$. The latter approximation, $D_{ij}(\lfloor \omega_{ij} \rfloor, \beta_{ij}) = \lfloor \omega_{ij} \rfloor + \lfloor \omega_{ij} \rfloor \beta_{ij}$, would be too small for larger values of β_{ij}, and this is why we propose $\tilde{D}_{ij}$ instead.

Although the approximating model

$$\eta_0 := \max_x \Big\{ \mathbb{E}_\omega[g_0(x,\omega)] : x \in X \Big\},$$

yielding the approximating solution $x_0 = (n_0, \beta_0)$, has integer first-stage decision variables, it is generally much easier to solve than the original problem (20) since this approximating model has a concave objective function. In Section 4.3 we use numerical experiments to analyze the quality of the solution $x_0 = (n_0, \beta_0)$. First, however, we derive an error bound for the concave approximation g_0, similar to that in Theorem 1.

4.2.2. Dual representation of g and g_0. To derive an upper bound on $\|G - G_0\|_\infty$ with $G_0(x) := \mathbb{E}_\omega[g_0(x,\omega)]$, $x \in X$, we first derive a dual representation of g and g_0. Here, we use the fact that, by Remark 4, g is a TU integer recourse function.

In analogous fashion to (8) and (9) in Section 2.1.3, we round down $D(\omega,\beta)$, relax the integrality constraints, and apply strong LP duality to obtain

$$g(x,\omega) = -q\beta + \min_{\mu,\lambda} \Big\{ \mu b(n) + \lambda \lfloor D(\omega,\beta) \rfloor : (\mu,\lambda) \in \Lambda \Big\}, \quad x \in X, \omega \in \mathbb{R}_+^{|E|}, \tag{26}$$

where $\Lambda := \{(\mu,\lambda) \in \mathbb{R}^{|V|} \times \mathbb{R}_+^{|E|} : \mu A + \lambda \geq r,\, \mu A \geq -c\}$. Similarly,

$$g_0(x,\omega) = -q\beta + \min_{\mu,\lambda} \Big\{ \mu b(n) + \lambda \tilde{D}(\omega,\beta) : (\mu,\lambda) \in \Lambda \Big\}, \quad x \in X, \omega \in \mathbb{R}_+^{|E|}. \tag{27}$$

Next, we derive, for a fixed $x \in X$, monotonicity properties of minimizers $(\mu(\omega), \lambda(\omega))$ and $(\tilde{\mu}(\omega), \tilde{\lambda}(\omega))$ of the optimization problems in $g(x,\omega)$ and $g_0(x,\omega)$, respectively, for every $\omega \in \mathbb{R}_+^{|E|}$. These properties are necessary to apply Proposition 1 to 'round up' $\lambda(\omega)$ and $\tilde{\lambda}(\omega)$ in the proof of Theorem 2.

LEMMA 2. *Let the directed acyclic graph $\mathcal{G}$ be t-connected. Consider the dual feasible region Λ defined as*

$$\Lambda = \Big\{ (\mu,\lambda) \in \mathbb{R}^{|V|} \times \mathbb{R}_+^{|E|} : \mu A + \lambda \geq r,\, \mu A \geq -c \Big\}$$

and let $x \in X$ be given. Let $H : \mathbb{R}_+^{|E|} \mapsto \mathbb{R}^{|E|}$ be a separable nonnegative function for which $H_{ij}(\omega_{ij})$ is non-decreasing in ω_{ij} for every $(i,j) \in E$, and let $\omega_{(ij)} \in \mathbb{R}^{|E|-1}$ denote ω without its ij-th component. Then, there exist minimizers $(\hat{\mu}(\omega), \hat{\lambda}(\omega))$ of

$$\min_{\mu,\lambda} \Big\{ \mu b(n) + \lambda H(\omega) : (\mu,\lambda) \in \Lambda \Big\} \tag{28}$$

such that for every $(i,j) \in E$ and $\omega_{(ij)} \geq 0$, the function $\hat{\lambda}_{ij}(\cdot|\omega_{(ij)}) : \mathbb{R}_+ \mapsto \mathbb{R}$ defined as $\hat{\lambda}_{ij}(\omega_{ij}|\omega_{(ij)}) = \hat{\lambda}_{ij}(\omega)$, satisfies *(i)* $\hat{\lambda}_{ij}(\cdot|\omega_{(ij)})$ *is monotone non-increasing, and*

(ii) $\hat{\lambda}_{ij}(\cdot|\omega_{(ij)}) \leq r_{ij} + c_{ij}$.

Proof: The proof of (i) is straightforward and similar to the proof of Lemma 11 in Romeijnders et al. (2016). Moreover, $\hat{\lambda}(\omega)$ is bounded since for every fixed $\hat{\mu}(\omega)$ an optimal solution $\hat{\lambda}_{ij}(\omega)$ of (28) takes values $\hat{\lambda}_{ij}(\omega) = (r_{ij} - (\hat{\mu}(\omega)A)_{ij})^+$. Here, we use the hypothesis that $H(\omega)$ is nonnegative and the fact that $\lambda \geq r - \mu A$ and $\lambda \geq 0$ for every $(\mu, \lambda) \in \Lambda$. Since $-\mu A \leq c$, we conclude that $\hat{\lambda}_{ij}(\omega) = (r_{ij} - (\hat{\mu}(\omega)A)_{ij})^+ \leq (r_{ij} + c_{ij})^+ = r_{ij} + c_{ij}$. Here, the last equality holds since $r_{ij}, c_{ij} > 0$, proving (ii). $\square$

Notice that for fixed $x \in X$, or more specifically for fixed investments, $\beta \geq 0$, both $\lfloor D(\omega, \beta) \rfloor$ and $\tilde{D}(\omega, \beta)$ satisfy the assumptions of $H(\omega)$ in Lemma 2. Moreover, the monotonicity result in (i) is one of the assumptions in Proposition 1 of Section 2.1.3. We can apply this proposition if for every $(i, j) \in E$, the function $\psi_{ij}(\omega_{ij}; \beta_{ij})$, defined as

$$\psi_{ij}(\omega_{ij}; \beta_{ij}) = \lfloor D_{ij}(\omega_{ij}, \beta_{ij}) \rfloor - \tilde{D}_{ij}(\omega_{ij}, \beta_{ij}), \tag{29}$$

is periodic in ω_{ij} for some period p_{ij} with zero mean $p_{ij}^{-1} \int_0^{p_{ij}} \psi_{ij}(t; \beta_{ij}) dt = 0$. If there are no investment effects, i.e., if $(i, j) \in E^0$, then this is trivially true since $\psi_{ij}(\omega_{ij}; \beta_{ij}) = 0$ for all $\omega_{ij} \geq 0$. The result also holds if the investment effects are additive, i.e., if $(i, j) \in E^+$, since in this case $\psi_{ij}(\omega_{ij}; \beta_{ij}) = \lfloor \omega_{ij} + \beta_{ij} \rfloor - \lfloor \omega_{ij} \rfloor - \beta_{ij} = \lfloor \omega_{ij} \rfloor_{-\beta_{ij}} - \lfloor \omega_{ij} \rfloor$, which is the same as $\bar{\varphi}_{z_i, \alpha_i}$ defined in Section 2.1.3 with $z_i := -\beta_{ij}$, $\alpha_i := 0$, and the round-up operators replaced by round-down operators. However, for multiplicative investment effects, i.e., for $(i, j) \in E^*$, the function $\psi_{ij}(\omega_{ij}; \beta_{ij})$, given for every $\omega_{ij} \geq 0$ by

$$\psi_{ij}(\omega_{ij}; \beta_{ij}) = \lfloor \omega_{ij}(1 + \beta_{ij}) \rfloor - \lfloor \omega_{ij} \rfloor - \omega_{ij}\beta_{ij}, \tag{30}$$

is periodic in ω_{ij} if and only if β_{ij} is rational, in which case its period is the least common multiple of 1 and $1/(1 + \beta_{ij})$. If β_{ij} is irrational, however, this least common multiple does not exist. This implies that for multiplicative investment effects the assumptions of Proposition 1 are not satisfied. We can circumvent this problem by decomposing $\psi_{ij}(\omega_{ij}; \beta_{ij})$ as the sum of two zero-mean periodic functions:

$$\psi_{ij}(\omega_{ij}; \beta_{ij}) = \Big(\lfloor \omega_{ij}(1 + \beta_{ij}) \rfloor - \omega_{ij}(1 + \beta_{ij}) + 1/2 \Big) + \Big(\omega_{ij} - 1/2 - \lfloor \omega_{ij} \rfloor \Big), \tag{31}$$

where the first and second periodic functions equal $\hat{\varphi}_{z_i}(\omega_{ij}(1 + \beta_{ij}))$ and $-\hat{\varphi}_{z_i}(\omega_{ij})$, respectively, with $\hat{\varphi}_{z_i}$ defined in Section 2.1.3 and in both cases with $z_i := 0$, the round-up operator replaced by a round-down operator, and the addition of 1.

4.2.3. Error bound. Now we are ready to derive an upper bound on $\|G - G_0\|_\infty$.

THEOREM 2. *Let $\mathcal{G} = (V,E)$ be a directed acyclic graph that is t-connected, and let $E^*, E^+, E^0 \subset E$ denote subsets of arcs with multiplicative, additive, and no investment effects, respectively. Let g denote the recourse function corresponding to the fleet allocation and routing problem defined for every $x \in X$ and $\omega \in \mathbb{R}_+^{|E|}$ as*

$$g(x,\omega) = -q\beta + \max_{y,z}\left\{ry - cz : Ay + Az = b(n),\, y \leq D(\omega,\beta),\, y,z \in \mathbb{Z}_+^{|E|}\right\},$$

and let g_0 denote its concave approximation defined in (25). Then, under the assumption that ω is a continuous random vector with joint pdf f and with independently distributed components, it holds for every $x \in X$, that

$$|G(x) - G_0(x)| \leq \sum_{(i,j)\in E^*\cup E^+} (r_{ij} + c_{ij}) h(|\Delta| f_{ij}),$$

where h is defined in (7), $G(x) := \mathbb{E}_\omega[g(x,\omega)]$ and $G_0(x) := \mathbb{E}_\omega[g_0(x,\omega)]$, $x \in X$, and $|\Delta| f_{ij}$ is the total variation of the marginal density function f_{ij}.

Proof: Let $x \in X$ and consider the dual representation of g in (26). Let $(\mu(\omega), \lambda(\omega))$ be minimizers of (28) with $H(\omega) := \lfloor D(\omega,\beta)\rfloor$, satisfying the properties of Lemma 2, so that $g(x,\omega) = -q\beta + \mu(\omega)b(n) + \lambda(\omega)\lfloor D(\omega,\beta)\rfloor$ for every $\omega \in \mathbb{R}_+^{|E|}$. Since $(\mu(\omega), \lambda(\omega))$ is feasible but not necessarily optimal for (28) with $H(\omega) := \tilde{D}(\omega,\beta)$, it follows from the dual representation of g_0 in (27) that $g_0(x,\omega) \leq -q\beta + \mu(\omega)b(n) + \lambda(\omega)\tilde{D}(\omega,\beta)$, and thus

$$G_0(x) - G(x) \leq \mathbb{E}_\omega\Big[\lambda(\omega)\Big(\tilde{D}(\omega,\beta) - \lfloor D(\omega,\beta)\rfloor\Big)\Big] = \sum_{(i,j)\in E} \mathbb{E}_\omega\Big[\lambda_{ij}(\omega)\Big(-\psi_{ij}(\omega_{ij};\beta_{ij})\Big)\Big],$$

where $\psi_{ij}(\omega_{ij};\beta_{ij})$ is defined in (29). Similarly, with $(\tilde{\mu}(\omega), \tilde{\lambda}(\omega))$ denoting minimizers of (28) with $H(\omega) := \tilde{D}(\omega,\beta)$, we have

$$G(x) - G_0(x) \leq \sum_{(i,j)\in E} \mathbb{E}_\omega\Big[\tilde{\lambda}_{ij}(\omega)\psi_{ij}(\omega_{ij};\beta_{ij})\Big]. \tag{32}$$

We will derive an upper bound on $G(x) - G_0(x)$; an upper bound for $G_0(x) - G(x)$ can be obtained in a similar way. We obtain this upper bound by separately bounding the individual terms, $\Psi_{ij}(\beta_{ij})$, in (32), defined for each $(i,j) \in E$ as $\Psi_{ij}(\beta_{ij}) = \mathbb{E}_\omega\Big[\tilde{\lambda}_{ij}(\omega)\psi_{ij}(\omega_{ij};\beta_{ij})\Big]$.

Obviously, if $(i,j) \in E^0$, then $\Psi_{ij}(\beta_{ij}) = 0$ since $\psi_{ij}(\omega_{ij};\beta_{ij}) = 0$ for all $\omega_{ij} \geq 0$. Moreover, if $(i,j) \in E^+$, then

$$\Psi_{ij}(\beta_{ij}) = \mathbb{E}_\omega\Big[\tilde{\lambda}_{ij}(\omega)\Big(\lfloor\omega_{ij}\rfloor_{-\beta_{ij}} - \lfloor\omega_{ij}\rfloor\Big)\Big] = \mathbb{E}_\omega\Big[\tilde{\lambda}_{ij}(\omega)\bar{\varphi}_{-\beta_{ij},0}(\omega_{ij})\Big],$$

where $\bar{\varphi}_{-\beta_{ij},0}$ is given in Definition 2. The second equality holds since we take the expectation with respect to a *continuously* distributed random vector ω, and thus it does not matter that by replacing the round-down operators by round-up operators we change the underlying difference function at countably many points $\omega_{ij} \in \{0, -\beta_{ij}\} + \mathbb{Z}$.

Writing the expectation as an integral and using the hypothesis that the components of ω are independent we have

$$\Psi_{ij}(\beta_{ij}) = \int_{\mathbb{R}^{|E|-1}}\left[\int_{\mathbb{R}}\tilde{\lambda}_{ij}(u_{ij}|u_{(ij)})\bar{\varphi}_{-\beta_{ij},0}(u_{ij})f_{ij}(u_{ij})du_{ij}\right]f_{(ij)}(u_{(ij)})du_{(ij)}.$$

Consider the inner integral for a fixed $u_{(ij)} \in \mathbb{R}^{|E|-1}$ and observe that $\tilde{\lambda}_{ij}(\cdot|u_{(ij)})$ is monotone non-increasing and bounded by $r_{ij}+c_{ij}$ according to Lemma 2. Moreover, $\bar{\varphi}_{-\beta_{ij},0}$ is periodic with zero mean value, so that all assumptions of Proposition 1 are satisfied. Applying this proposition to the inner integral yields

$$\begin{aligned}
\Psi_{ij}(\beta_{ij}) &\leq \int_{\mathbb{R}^{|E|-1}}(r_{ij}+c_{ij})M(\bar{\varphi}_{-\beta_{ij},0},|\Delta|f_{ij})f_{(ij)}(u_{(ij)})du_{(ij)} \\
&= (r_{ij}+c_{ij})M(\bar{\varphi}_{-\beta_{ij},0},|\Delta|f_{ij}) \\
&\leq (r_{ij}+c_{ij})h(|\Delta|f_{ij}),
\end{aligned}$$

where M is defined in (12) and the last inequality follows from (14).

It remains to show that $\Psi_{ij}(\beta_{ij}) \leq (r_{ij}+c_{ij})h(|\Delta|f_{ij})$ for $(i,j) \in E^*$. In this case, using (31), we have

$$\begin{aligned}
\Psi_{ij}(\beta_{ij}) &= \mathbb{E}_\omega\Big[\tilde{\lambda}_{ij}(\omega)\Big(\lfloor\omega_{ij}(1+\beta_{ij})\rfloor - \lfloor\omega_{ij}\rfloor - \omega_{ij}\beta_{ij}\Big)\Big] \\
&= \mathbb{E}_\omega\Big[\tilde{\lambda}_{ij}(\omega)\Big(-\hat{\varphi}_0(\omega_{ij})\Big)\Big] + \mathbb{E}_\omega\Big[\tilde{\lambda}_{ij}(\omega)\hat{\varphi}_0\Big(\omega_{ij}(1+\beta_{ij})\Big)\Big],
\end{aligned}$$

with $\hat{\varphi}_0$ given in Definition 2, and where again the second equality holds even though we replaced the round-down operators by round-up operators. Since $\hat{\varphi}_0$ is periodic with zero mean value, we can apply Proposition 1 twice, in a similar way as for $(i,j) \in E^+$, to obtain

$$\Psi_{ij}(\beta_{ij}) \leq (r_{ij}+c_{ij})M\Big(-\hat{\varphi}_0,|\Delta|f_{ij}\Big) + (r_{ij}+c_{ij})M\Big(\hat{\varphi}_0,\frac{|\Delta|f_{ij}}{1+\beta_{ij}}\Big). \tag{33}$$

Here we use Lemma 1(iii) of Romeijnders et al. (2016) for the second term, recognizing that pdf f'_{ij} of $\omega'_{ij} = \omega_{ij}(1+\beta_{ij})$ has total variation $|\Delta| f'_{ij} = (1+\beta_{ij})^{-1}|\Delta| f_{ij}$. Inserting the expressions of (14) for $M(\hat{\varphi}_0, B)$ and $M(-\hat{\varphi}_0, B)$ into (33), we have

$$\begin{aligned}\Psi_{ij}(\beta_{ij}) &\leq \frac{1}{2}(r_{ij}+c_{ij})h(|\Delta| f_{ij}) + \frac{1}{2}(r_{ij}+c_{ij})h\Big(\frac{|\Delta| f_{ij}}{1+\beta_{ij}}\Big) \\ &\leq (r_{ij}+c_{ij})h(|\Delta| f_{ij}),\end{aligned}$$

where the second inequality holds because h is increasing and $\beta_{ij} \geq 0$.

Substituting the bounds on $\Psi_{ij}(\beta_{ij})$ into (32) yields

$$G(x) - G_0(x) \leq \sum_{(i,j)\in E^* \cup E^+} (r_{ij}+c_{ij})h(|\Delta| f_{ij}).$$

As already mentioned, the same upper bound can be obtained for $G_0(x) - G(x)$ using a similar line of reasoning. $\square$

4.3. Computational study

We use numerical experiments to evaluate the actual performance of the concave approximation, defined in (25), of the fleet allocation and routing problem. All experiments are carried out on the graph instance of Donohue and Birge (1995), given in Figure 2, and the cost and reward parameters, r and c, are also taken from this reference. The computational results we report use GAMS 24.2.1 and IBM ILOG CPLEX 12.6 to solve the mixed-integer linear programs (MILPs) on a Dell Poweredge 2950 computer with two dual-core Intel (Xenon) 3.73 GHz Xeon processors and 24 GB of shared memory running Ubuntu Linux.

4.3.1. Experimental design. We assume that all random variables, ω_{ij}, are independently distributed and follow a truncated normal distribution; i.e., $\omega_{ij} := [\bar{\omega}_{ij} | \bar{\omega}_{ij} \geq 0]$, where $\bar{\omega}_{ij} \sim N(\mu_{ij}, \sigma^2)$. The mean, μ_{ij}, is the same as in Donohue and Birge (1995) and differs per arc (i,j), whereas the standard deviation σ is the same for each arc, but varies over the experiments. We also vary the investment cost parameters q by defining $q = \kappa_q(r+c)$ and selecting different values for the scalar inflation factor, κ_q.

We let $\kappa_q \in \{0.2, 0.5, 0.8\}$ and $\sigma \in \{0.1, 1, 10\}$ so that the values of σ correspond to low, medium, and high variability. Moreover, we consider two settings, one with additive investment effects and one with multiplicative investment effects. In the first case $E^+ = \{(1,8),(4,9),(7,13),(11,16),(14,17)\}$ and in the second case E^* is the same arc set. (Again, see Figure 2.) For other experiments with different arc sets we obtained similar results.

4.3.2. Numerical results. We evaluate the performance of the α-approximation with $\alpha = 0$, defined in (25), using the same approach as in Section 3.2 for the integer newsvendor problem. Here, we apply the MRP with $n = 50$ instead of $n = 1000$, because of the increased computational effort required to solve the deterministic equivalent MILP. In fact, for each MILP we stop the branch-and-bound procedure at either a relative tolerance of 0.001% or after five minutes of computation time, which ever occurs first, and we let x_n^{i*} be the best integer solution obtained.

To obtain the approximating solution, x_0, we solve the approximating problem using a sample average approximation with a sample size of 250. The deterministic equivalent problem of this approximation is also a MILP, but it is easier to solve because the approximating model has integer variables in the first stage only. Finally, to obtain the sampling solution, x^S, we compare the solutions x_n^{i*}, $i = 1, \ldots, N_r$, using an out-of-sample estimation with a sample of size 1000. Note that we do not consider the average of the solutions x_n^{i*} because this average is not necessarily feasible.

We report the same performance measures $\rho_1(x_0), \rho_2(x_0), \rho_2(x^S)$, and $\rho_3(x_0, x^S)$ as in Section 3.2. However, here the denominator in $\rho_1(x_0)$ is Theorem 2's error bound. Thus, $\rho_1(x_0)$ compares the MRP optimality gap with the total variation error bound; i.e., it estimates the sharpness of the total variation error bound. Moreover, $\rho_2(x_0)$ and $\rho_2(x^S)$ give the width of the MRP's confidence interval on the respective optimality gaps as a percentage of the estimate of the model's optimal objective function value; i.e., they estimate the quality of the approximating solutions x_0 and x^S. Furthermore, $\rho_3(x_0, x^S)$ estimates the difference in the objective function values of the approximating solution and the sampling solution. To estimate $G(x_0) - G(x^S)$ we use a sample of size $10,000$. Even though the MILPs are not solved to optimality, we use x_n^{i*} in Step 1(iii) of the MRP and its objective function value η_n^{i*} in the denominators of ρ_2 and ρ_3. As a result, the values of $\rho_1(x_0)$, $\rho_2(x_0)$, and $\rho_2(x^S)$ may be too small when the (deterministic) MILP optimality gap is not small enough. For this reason, we also report Γ, a bound on the (deterministic) optimality gap of these MILPs as a percentage of the objective function value. Summing Γ and ρ_2 has the effect of replacing the objective function value of x_n^{i*} with the MILP relaxation bound in ρ_2's numerator.

Table 5's results for additive investment effects are similar to Section 3.2's results for the integer newsvendor problem. The concave approximation is good in case of medium

Table 5 Numerical results for the fleet allocation and routing problem. The effect of investments are additive, $E^+ = \{(1,8),(4,9),(7,13),(11,16),(14,17)\}$, and the ρ- and Γ-values are reported as percentages. Because we are maximizing, positive values for ρ_3 mean that x_0 outperforms x^S and negative values mean the opposite.

Exp.	σ	κ_q	Γ	$\rho_1(x_0)$	$\rho_2(x_0)$	$\rho_2(x^S)$	$\rho_3(x_0,x^S)$
1	0.1	0.2	0.00	5.4	0.69	0.05	-0.63
2	0.1	0.5	0.03	15.2	2.51	0.12	-2.39
3	0.1	0.8	0.15	23.1	4.86	0.21	-4.63
4	1	0.2	0.07	11.6	0.20	0.21	0.02
5	1	0.5	0.69	17.6	0.38	0.39	0.03
6	1	0.8	1.55	17.0	0.44	0.53	0.12
7	10	0.2	0.00	34.6	0.06	0.07	0.003
8	10	0.5	0.00	22.3	0.04	0.05	0.003
9	10	0.8	0.00	30.8	0.05	0.06	0.008

and high variability ($\rho_2(x_0)$ is smaller than 1%, and also the values of Γ are fairly small) and the approximation is worse in the case of low variability. Indeed, for Experiment 3 the value of $\rho_2(x_0)$ is almost 5%. Although not visible in the table, we observe that in line with the total variation error bound, the width of the MRP's confidence interval on the optimality gap decreases as σ increases.

Moreover, the actual performance of the concave approximation may be significantly better than implied by the (worst-case) error bound of Theorem 2. In particular, for $\sigma = 1$ the values of $\rho_1(x_0)$ are smallest and below 20%. Interestingly, the values of $\rho_1(x_0)$ are higher for $\sigma = 10$. This is caused by the fact that the total variation error bound (i.e., the denominator of $\rho_1(x_0)$) decreases as σ increases, whereas the MRP optimality gap (i.e., the numerator of $\rho_1(x_0)$) slightly increases due to the increased variability in the model. These effects offset the decrease in the MRP optimality gap caused by the fact that the approximating solution x_0 is bad for $\sigma = 0.1$ and good for $\sigma = 10$.

For medium and high variability, $\rho_3(x_0, x^S)$ indicates that the concave approximation is as good as the sampling solution. Keeping in mind that obtaining the sampling solution x^S requires much more computational effort, we prefer to use the concave approximation under these circumstances. (The typical time to compute x_0 with a sample size of 250 is one second, whereas the time to compute a single x_n^{i*} with a sample size of 50 often exceeds five minutes.) For low variability, however, the sampling solution is better (as can be observed from the values of $\rho_3(x_0, x^S)$), and is in fact close to optimal (since the values of $\rho_2(x^S)$ and Γ are small). Moreover, for low variability the sampling method typically gives good solutions even if the sample size is small, so that the sampling method can be carried out within reasonable time limits. Thus, in some sense the concave approximation and the sampling method can be considered as complementary approaches: the concave approximation is preferred in case of medium and high variability, and the sampling method in case of low variability.

For multiplicative investment effects we obtain similar results; see Table 6. The main difference is that the performance of the concave approximation is also good in the low variability case. This may be caused by the fact that for additive investment effects and $\sigma = 0.1$, the demands, $\tilde{D}_{ij}(\omega_{ij}, \beta_{ij}) = \lfloor \omega_{ij} \rfloor + \beta_{ij}$, in the approximating model are almost deterministic, whereas for multiplicative investment effects the variability in the demands, $\tilde{D}_{ij}(\omega_{ij}, \beta_{ij}) = \lfloor \omega_{ij} \rfloor + \omega_{ij}\beta_{ij}$, is larger.

Table 6 Numerical results for the fleet allocation and routing problem. The effect of investments are multiplicative, $E^* = \{(1,8),(4,9),(7,13),(11,16),(14,17)\}$, and the ρ- and Γ-values are reported as percentages. Because we are maximizing, positive values for ρ_3 mean that x_0 outperforms x^S and negative values mean the opposite.

Exp.	σ	κ_q	Γ	$\rho_1(x_0)$	$\rho_2(x_0)$	$\rho_2(x^S)$	$\rho_3(x_0, x^S)$
1	0.1	0.2	0.00	1.28	0.15	0.06	-0.09
2	0.1	0.5	0.25	1.56	0.20	0.12	-0.09
3	0.1	0.8	0.60	5.10	0.75	0.18	-0.55
4	1	0.2	0.02	15.3	0.25	0.28	-0.01
5	1	0.5	0.59	20.2	0.38	0.43	0.04
6	1	0.8	1.22	21.4	0.45	0.47	-0.01
7	10	0.2	0.00	73.1	0.12	0.13	0.006
8	10	0.5	0.00	51.1	0.08	0.09	0.009
9	10	0.8	0.00	40.6	0.07	0.07	0.009

5. Investment in stochastic activity networks

This section discusses an investment problem in a stochastic activity network, based on network instances from Elmaghraby (1977). As for the fleet allocation and routing problem it may be viewed as a two-stage totally unimodular integer recourse model. The error bound in Theorem 1 cannot be applied directly because there may be perfect dependencies in the right-hand side random vector due to the special structure of the model. Using properties of the optimal dual variables of the second-stage problem we circumvent this difficulty and derive an error bound for a convex approximation of this investment problem.

5.1. Problem definition and model formulation

Section 5.1.1 defines the problem. Then, Section 5.1.2 formulates the stochastic integer program, and Section 5.1.3 discusses properties of the model.

5.1.1. Problem definition. Consider a project in which several activities, indexed by $i = 1, \dots, N$, need to be carried out, but where the durations D_i of these activities (measured in days) are random. Moreover, there are precedence constraints on the activities, represented by the arcs E of a directed acyclic graph $\mathcal{G} = (V, E)$. That is, for every $(i, j) \in E$, activity j cannot start before activity i is completed. In addition, we assume that activities can only start at the beginning of a day.

Before the start of the project there is a budget B available to invest in the activities $i = 1, \dots, N$. Investing x_i in activity i, at unit cost c_i, reduces its duration D_i. We assume that the effects of the investment may be additive ($D_i(\omega, x) = \omega_i - x_i$) or multiplicative ($D_i(\omega, x) = \omega_i(1 - x_i)$), and that the duration $D_i(\omega, x) \geq 0$ for every possible realization of ω and investment x. The problem is to invest $x \geq 0$ in the activities subject to the budget constraint $cx \leq B$ to minimize the expected completion time of the project.

5.1.2. Model formulation: two-stage recourse. The problem can be formulated as a two-stage integer recourse model, where in the first stage we determine the investments x, and where in the second stage we determine the start times $t \in \mathbb{Z}_+^N$ (in days) of the activities given the durations. We assume that activity N is a dummy activity with duration 0 that can only start if all other activities are finished. The two-stage integer recourse model for this problem is thus given by

$$\eta^* := \min_x \left\{ \mathbb{E}_\omega \Big[g(x, \omega) \Big] : x \in X \right\},$$

where for every $x \in X$ and $\omega \in \Omega$,

$$g(x, \omega) := \min_t \left\{ t_N : t_j - t_i \geq D_i(\omega, x) \text{ for all } (i, j) \in E, \quad t \in \mathbb{Z}_+^N \right\}$$

and

$$X = \left\{ x \in \mathbb{R}_+^N : cx \leq B, \, x \leq b \right\}.$$

Here, we assume that x is bounded by b such that $D(\omega, x) \geq 0$ for all $x \in X$ and $\omega \in \Omega$, and that the graph $\mathcal{G}$ is N-connected. Moreover, the first constraints in the problem defining g represent the precedence relations between the activities, and the integrality constraints on t model the idea that each activity starts at the beginning of a day.

Remark 5. We could instead assume that $t_N \in \mathbb{R}_+$. Also in this case we can derive a convex approximation and corresponding error bound. However, we prefer $t_N \in \mathbb{Z}_+$ for ease of exposition.

Let A denote the node-arc incidence matrix of $\mathcal{G}$, where the rows of A correspond to the arcs of $\mathcal{G}$ and the columns of A to the nodes of $\mathcal{G}$. The row of A corresponding to $(i,j) \in E$ has one entry equal to -1 in column i and $+1$ in column j, and the remaining entries are equal to zero. Moreover, we define $d(\omega,x) \in \mathbb{R}_+^{|E|}$ as $d_{ij}(\omega,x) = D_i(\omega,x)$ for every $(i,j) \in E$. Then, the precedence constraints can be written as $At \geq d(\omega,x)$ and thus for every $x \in X$ and $\omega \in \Omega$,

$$g(x,\omega) = \min_t \left\{ e_N t : At \geq d(\omega,x),\, t \in \mathbb{Z}_+^N \right\}, \tag{34}$$

where e_N is the N-th unit vector.

Remark 6. Observe that some of the components of the right-hand side random vector $d(\omega,x)$ may be identical, so that we will not be able to apply Theorem 1 directly to derive an error bound for the convex approximation to be defined in Section 5.2.1.

Remark 7. Note that in the notation of Section 4, the recourse matrix in (34) is $-A^\top$.

5.1.3. Properties of the recourse function g. Next we discuss properties of the recourse function g. We show that the recourse is complete and sufficiently expensive.

Lemma 3. *Let $\mathcal{G} = (V,E)$ be a directed acyclic graph, let g be the recourse function defined in (34), and let $G(x) := \mathbb{E}_\omega[g(x,\omega)]$, $x \in X$. Then, the recourse is complete and sufficiently expensive; that is,*

$$g(x,\omega) \text{ is finite for every } x \in \mathbb{R}^N \text{ and } \omega \in \mathbb{R}^N.$$

Moreover, if $\mathbb{E}_\omega[(\omega_i)^+] < +\infty$ for all $i = 1,\ldots,N$, then

$$G(x) \text{ is finite for all } x \in \mathbb{R}^N.$$

Proof: Since $t_N \geq 0$ it follows immediately that for every $x \in \mathbb{R}^N$ and $\omega \in \mathbb{R}^N$, $g(x,\omega) \geq 0$. Moreover, since $\mathcal{G}$ is acyclic it is always possible to find a feasible solution $t \in \mathbb{Z}_+^N$ for a given $x \in \mathbb{R}^N$ and $\omega \in \mathbb{R}^N$. Thus, the recourse is complete and sufficiently expensive. This implies the finiteness of G under the assumption $\mathbb{E}_\omega[(\omega_i)^+] < +\infty$ for all $i = 1,\ldots,N$. □

Remark 8. Since the recourse matrix A is a node-arc incidence matrix, it follows immediately that A is TU.

5.2. Convex approximation

We derive a convex approximation of g using the same type of approximations as in Section 2.1.1. Here, we prefer to use a shifted LP-relaxation approximation since in Section 3 we observed it outperforms the α-approximations. In Section 5.2.1 we define the convex approximation $\hat{g}$, in Section 5.2.2 we present a dual representation of g and $\hat{g}$ and we derive properties of the optimal dual variables, and in Section 5.2.3 we derive an error bound for the convex approximation.

5.2.1. Definition of the convex approximation $\hat{g}$. Similarly as in Section 2.1.1 we obtain a convex approximation $\hat{g}$ of g by simultaneously relaxing the integrality constraints and replacing the right-hand side random vector $d(\omega,x)$ by $\hat{d}(\omega,x)$. We define the latter as follows. For every $(i,j)\in E$,

$$\hat{d}_{ij}(\omega,x)=\begin{cases} d_{ij}(\omega,x)+1/2, & \text{if } i\in V^+\cup V^* \\ \lceil d_{ij}(\omega,x)\rceil, & \text{if } i\in V^0. \end{cases}$$

Here, V^*, V^+, and V^0 partition V, which represents the activities, into subsets with multiplicative, additive, and no investment effects, respectively. Observe that if for some activity i, there is no investment possible, then we do not add 1/2 to the right-hand sides $d_{ij}(\omega,x)$ but instead we round up $d_{ij}(\omega,x)$, which in that case does not depend on x. This yields a smaller error bound, as we will see in Section 5.2.3. The convex approximation is for every $x\in X$ and $\omega\in\Omega$ given by

$$\hat{g}(x,\omega)=\min_t\left\{e_N t : At\geq \hat{d}(\omega,x),\, t\in\mathbb{R}^N_+\right\}. \tag{35}$$

5.2.2. Dual representations of g and $\hat{g}$. To derive an upper bound on $|G(x)-\hat{G}(x)|$ for every $x\in X$, where $\hat{G}(x)=\mathbb{E}_\omega[\hat{g}(x,\omega)]$, we first derive dual representations of g and $\hat{g}$. Both the dual of g and $\hat{g}$ can be interpreted as a longest-path problem in a directed acyclic graph, and the optimal dual variables indicate a critical path in the graph.

In analogous fashion to (8) and (9) in Section 2.1.3, we round up the right-hand side $d(\omega,x)$, relax the integrality constraints, and apply strong LP duality to obtain

$$g(x,\omega)=\max_\pi\left\{\pi\lceil d(\omega,x)\rceil : \pi A\leq e_N,\, \pi\in\mathbb{R}^{|E|}_+\right\},$$

or equivalently,

$$g(x,\omega)=\max_{\pi\geq 0}\left\{\pi\lceil d(\omega,x)\rceil : \begin{array}{rl} \sum_{(k,i)\in E}\pi_{ki}-\sum_{(i,j)\in E}\pi_{ij}\leq 0, & \text{if } i\neq N, \\ \sum_{(k,i)\in E}\pi_{ki}\leq 1, & \text{if } i=N. \end{array}\right\}$$

Since by Lemma 3 the recourse is complete and sufficiently expensive, it follows that the dual feasible region is non-empty and bounded, and thus it can be characterized by finitely many vertices π^k, $k = 1, \dots, K'$. Moreover, when $\lceil d(\omega,x) \rceil \geq 0$ we can restrict attention to dual vertices that correspond to a maximal path P^k in the graph $\mathcal{G}$ with $\pi^k_{ij} = 1$ if arc $(i,j) \in E$ is on the path P^k and $\pi^k_{ij} = 0$, otherwise. These paths $P^k, k = 1, \dots, K$, are maximal in the sense that they are not subpaths of any other directed path in $\mathcal{G}$. In any case, we can rewrite g for every $x \in X$ and $\omega \in \Omega$ as

$$g(x,\omega) = \max_{k=1,\dots,K} \pi^k \lceil d(\omega,x) \rceil .$$

For every $k = 1, \dots, K$, we introduce indicator variables Π^k_i, $i = 1, \dots, N$, that indicate whether node i is on path P^k ($\Pi^k_i = 1$) or not ($\Pi^k_i = 0$). Since the directed graph $\mathcal{G}$ is acyclic we can derive the value Π^k directly from π^k: for every $k = 1, \dots, K$, we have

$$\Pi^k_i = \sum_{(i,j)\in E} \pi^k_{ij}, \quad i = 1, \dots, N. \tag{36}$$

Observe that even though node N is on every path P^k, the indicator variable $\Pi^k_N = 0$ since node N has no outgoing arcs. However, whether node N is on path P^k is irrelevant since its duration is zero. Using variables $\Pi^k := (\Pi^k_1, \dots, \Pi^k_N)$ we can rewrite the dual objective:

$$\begin{aligned}\pi^k \lceil d(\omega,x) \rceil &= \sum_{(i,j)\in E} \pi^k_{ij} \lceil d_{ij}(\omega,x) \rceil = \sum_{i=1}^{N} \Big\{ \sum_{(i,j)\in E} \pi^k_{ij} \lceil D_i(\omega,x) \rceil \Big\} = \sum_{i=1}^{N} \Pi^k_i \lceil D_i(\omega,x) \rceil \\ &= \Pi^k \lceil D(\omega,x) \rceil ,\end{aligned}$$

yielding

$$g(x,\omega) = \max_{k=1,\dots,K} \Pi^k \lceil D(\omega,x) \rceil . \tag{37}$$

For the convex approximation $\hat{g}$ we can derive a similar dual representation. We have

$$\begin{aligned}\hat{g}(x,\omega) &= \max_{k=1,\dots,K} \pi^k \hat{d}(\omega,x) \\ &= \max_{k=1,\dots,K} \Pi^k \hat{D}(\omega,x),\end{aligned} \tag{38}$$

where

$$\hat{D}_i(\omega,x) = \begin{cases} D_i(\omega,x) + 1/2, & \text{if } i \in V^+ \cup V^* \\ \lceil D_i(\omega,x) \rceil , & \text{if } i \in V^0. \end{cases}$$

Next, for arbitrary but fixed $x \in X$, we derive monotonicity properties of maximizers $\Pi(\omega)$ and $\hat{\Pi}(\omega)$ of the optimization problems in (37) and (38), respectively. These properties are necessary to apply Proposition 1, allowing us to obtain the desired error bound.

LEMMA 4. *Let the directed acyclic graph $\mathcal{G}=(V,E)$ be N-connected. Consider the maximal paths $P^k, k=1,\dots,K$, of $\mathcal{G}$ with corresponding indicator variables Π^k defined in (36). Let $H:\mathbb{R}^N_+\mapsto\mathbb{R}^N$ be a separable nonnegative function for which $H_i(\omega_i)$ is non-decreasing in ω_i for every $i=1,\dots,N$, and let $\omega_{(i)}$ denote ω without its i-th component. Then, there exist maximizers $\Pi(\omega)$ of*

$$\max_{k=1,\dots,K}\Pi^k H(\omega)$$

such that for every $i=1,\dots,N$ and $\omega_{(i)}\geq 0$, the function $\Pi_i(\cdot|\omega_{(i)}):\mathbb{R}_+\mapsto\mathbb{R}$ defined as $\Pi_i(\omega_i|\omega_{(i)})=\Pi_i(\omega)$ satisfies

(i) $\Pi_i(\cdot|\omega_{(i)})$ *is monotone non-decreasing, and*

(ii) $\Pi_i(\cdot|\omega_{(i)})\leq 1$.

Proof: The proof of (i) is straightforward and similar to the proof of Lemma 11 in Romeijnders et al. (2016). It can be interpreted as follows. If node i is on the critical path and we only increase the arc lengths corresponding to the outgoing arcs from node i, then node i will stay on the critical path. Moreover, (ii) holds since Π_i is an indicator variable. $\square$

5.2.3. Error bound. Now we are ready to derive an upper bound on $|G(x)-\hat{G}(x)|$ for all $x\in X$.

THEOREM 3. *Let $\mathcal{G}=(V,E)$ be a directed acyclic graph that is N-connected, and let $V^*,V^+,V^0\subset V$ denote subsets of nodes, i.e., activities, with multiplicative, additive, and no investment effects, respectively. Let g denote the recourse function corresponding to the stochastic activity network investment problem defined for every $x\in X$ and $\omega\in\mathbb{R}^N_+$ as*

$$g(x,\omega)=\min_t\left\{e_N t: At\geq d(\omega,x),\ t\in\mathbb{Z}^N_+\right\},$$

and let $\hat{g}$ denote its convex approximation defined in (35). Moreover, define $G(x):=\mathbb{E}_\omega[g(x,\omega)]$ and $\hat{G}(x):=\mathbb{E}_\omega[\hat{g}(x,\omega)], x\in X$. Then, under the assumption that ω is a continuous random vector with joint pdf f and with independently distributed components, and with $\mathbb{E}_\omega[(\omega_i)^+]<+\infty$ for all $i=1,\dots,N$, it holds for every $x\in X$, that

$$|G(x)-\hat{G}(x)|\leq\frac{1}{2}\sum_{i\in V^+}h(|\Delta|f_i)+\frac{1}{2}\sum_{i\in V^*}h\left(\frac{|\Delta|f_i}{1-b_i}\right),$$

where h is defined in (7), $|\Delta|f_i$ is the total variation of the marginal density function f_i, and b_i is an upper bound for x_i such that $d(\omega,x)\geq 0$ for all $x\in X$ and nonnegative ω.

Proof: Let $x \in X$ be given and consider the dual representations of g and $\hat{g}$ as defined in (37) and (38) given by

$$g(x,\omega) = \max_{k=1,\dots,K} \Pi^k \lceil D(\omega,x) \rceil$$

and

$$\hat{g}(x,\omega) = \max_{k=1,\dots,K} \Pi^k \hat{D}(\omega,x).$$

For every $\omega \in \mathbb{R}^N_+$, let $\Pi(\omega)$ denote an optimal solution corresponding to $g(x,\omega)$. Since $\Pi(\omega)$ is not necessarily a maximizer for $\hat{g}(x,\omega)$, it follows that $\hat{g}(x,\omega) \leq \Pi(\omega)\hat{D}(\omega,x)$, and thus

$$\begin{aligned} G(x) - \hat{G}(x) &\leq \mathbb{E}_\omega\Big[\Pi(\omega)\Big(\lceil D(\omega,x)\rceil - \hat{D}(\omega,x)\Big)\Big] \\ &= \sum_{i=1}^N \mathbb{E}_\omega\Big[\Pi_i(\omega)\Big(\lceil D_i(\omega,x)\rceil - \hat{D}_i(\omega,x)\Big)\Big]. \end{aligned} \tag{39}$$

Similarly, there exist optimal solutions $\hat{\Pi}(\omega)$ in $\hat{g}(x,\omega)$ so that

$$\hat{G}(x) - G(x) \leq \sum_{i=1}^N \mathbb{E}_\omega\Big[\hat{\Pi}_i(\omega)\Big(\hat{D}_i(\omega,x) - \lceil D_i(\omega,x)\rceil\Big)\Big].$$

Analogously to the proof of Theorem 1 we only derive an upper bound for $G(x) - \hat{G}(x)$. Moreover, we again separately bound the individual terms, $\Psi_i(x_i)$, in (39) defined for every $i = 1,\dots,N$, as

$$\Psi_i(x_i) = \mathbb{E}_\omega\Big[\Pi_i(\omega)\psi_i(\omega_i;x_i)\Big],$$

where

$$\psi_i(\omega_i;x_i) = \lceil D_i(\omega,x)\rceil - \hat{D}_i(\omega,x).$$

Since $\lceil D(\omega,x)\rceil$ is a separable non-decreasing function in ω it follows from Lemma 4 that $\Pi_i(\omega)$ in non-decreasing in ω_i for every $\omega_{(i)} \geq 0$. This is one of the conditions of Proposition 1. The other condition is that $\psi_i(\omega_i;x_i)$ is periodic in ω_i, which we show next.

For $i \in V^0$ we have $\psi_i(\omega_i;x_i) = \lceil\omega_i\rceil - \lceil\omega_i\rceil = 0$, which is of course periodic in ω_i. Moreover, for $i \in V^+$, we have

$$\psi_i(\omega_i;x_i) = \lceil\omega_i - x_i\rceil - (\omega_i + 1/2 - x_i) = \hat{\varphi}_{x_i}(\omega_i),$$

where $\hat{\varphi}_{x_i}$ is specified in Definition 2. Similarly, for $i \in V^*$,

$$\psi_i(\omega_i;x_i) = \lceil\omega_i(1-x_i)\rceil - \Big(\omega_i(1-x_i) + 1/2\Big) = \hat{\varphi}_0\Big(\omega_i(1-x_i)\Big).$$

Hence, using Proposition 1 we have $\Psi_i(x_i) = 0$ if $i \in V^0$. If $i \in V^+$, then

$$\Psi_i(x_i) \leq M(\hat{\varphi}_{x_i}, |\Delta| f_i) = \frac{1}{2} h(|\Delta| f_i),$$

and if $i \in V^*$, then

$$\Psi_i(x_i) \leq M\left(\hat{\varphi}_0, \frac{|\Delta| f_i}{1 - x_i}\right) = \frac{1}{2} h\left(\frac{|\Delta| f_i}{1 - x_i}\right) \leq \frac{1}{2} h\left(\frac{|\Delta| f_i}{1 - b_i}\right),$$

where the last inequality holds since h is increasing and $x_i \leq b_i$. Combining all results yields the error bound. $\square$

Remark 9. To ensure that $D(\omega, x) \geq 0$ for all $x \in X$ and nonnegative ω, we require that $b_i \geq 0$ for all $i \in V$. Moreover, for every $i \in V^*$, we require $b_i < 1$.

Remark 10. In Theorem 3, we assume that the components of ω are independently distributed. In case these components are dependent a similar error bound may be derived depending on the total variations of the one-dimensional conditional density functions; see Romeijnders et al. (2015, 2016).

5.3. Computational study

We use numerical experiments to evaluate the actual performance of the convex approximation, defined in (35), of the stochastic activity network investment problem. We consider a single stochastic activity network given in Elmaghraby (1977) and also used in Avramdidis et al. (1991). Figure 3 depicts the precedence relations between the activities. The computational results we report use GUROBI 6.0.4 and Python 2.7 to solve the MILPs.

5.3.1. Experimental design. We assume that all random variables, ω_i, are independently distributed and follow a truncated normal distribution. The mean, μ_i, is the same as in Elmaghraby (1977) and Avramdidis et al. (1991) and differs per activity i, whereas the standard deviation σ is the same for each arc, but varies over the experiments. Furthermore, we also vary the budget B and the upper bound, b, on the investments; the per unit investment costs are $c_i = 1$ for every activity and every experiment.

We consider two settings, one with additive investment effects and one with multiplicative investment effects. In the first case, $V^+ = \{1, 6, 7, 12, 13\}$ and in the second case V^* denotes the same set of nodes. In both settings we use $\sigma \in \{0.1, 0.3, 1, 3, 10\}$ but we use different values for B and b since for multiplicative investment effects $b_i < 1$ is required.

In the first setting with additive investment effects we use $B \in \{3, 8\}$ and $b_i = (\mu_i - 2\sigma)^+$ for each activity $i = 1, \ldots, N$. Under mean times for each activity, the project's duration is

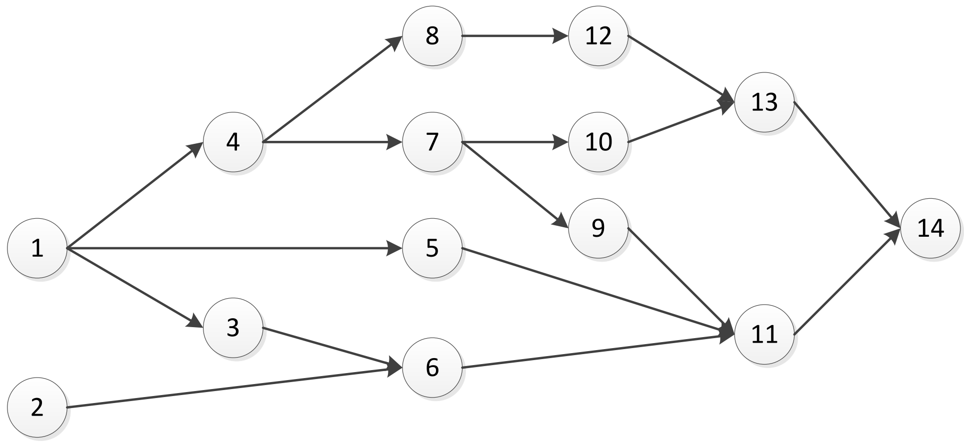

Figure 3 **The graph $\mathcal{G} = (V, E)$ used for the numerical experiments. Nodes 1-13 are activity nodes, node 14 represents a dummy activity, and the directed arcs represent precedence relations.**

approximately 40. The durations ω_i are truncated so that $\omega_i \geq b_i$ for all $\omega \in \Omega$. In the second setting with multiplicative investment effects we use $B = 1$ and $b \in \{0.3, 0.7\}$, meaning that either $b_i = 0.3$ for each activity i or $b_i = 0.7$ for each activity. In this case, the durations ω_i are truncated so that $\omega_i \geq 0$ for all $\omega \in \Omega$.

5.3.2. Numerical results. We evaluate the performance of the convex approximation, defined in (35), using the same approach as in Section 4.3.2 for the fleet allocation and routing problem. That is, we apply the MRP with $n = 50$ and we obtain an approximating solution $\hat{x}$ by solving a sample average approximation with a sample size of 250. We report the performance measures $\rho_1(\hat{x}), \rho_2(\hat{x}), \rho_2(x^S)$, and $\rho_3(\hat{x}, x^S)$ of Section 2.2.2, where the denominator in $\rho_1(\hat{x})$ is Theorem 3's error bound.

Table 7's results for additive investment effects are similar to those for the integer newsvendor problem in Section 3 and the fleet allocation and routing problem in Section 4. That is, solutions from the convex approximation are of good quality for medium and high variability (the values of $\rho_2(\hat{x})$ are below 0.5%) but not so good for low variability. In this case, the sampling solution x^S is better (e.g., $\rho_3(\hat{x}, x^S) = -1.57$ for $\sigma = 0.1$ and $B = 3$), whereas x^S is comparable to the approximating solution $\hat{x}$ in other cases. The value of $\rho_1(\hat{x})$ is high for low and high variability. In the first case, this is because the quality of the approximating solution $\hat{x}$ is not good, whereas in the second case this is because the total variation error bound decreases in σ while the MRP optimality gap stays approximately

Table 7 Numerical results for the stochastic activity network investment problem. The effect of investments are additive, $V^+ = \{1, 6, 7, 12, 13\}$, and the ρ-values are reported as percentages. Positive values for ρ_3 mean that $\hat{x}$ outperforms x^S and negative values mean the opposite.

Exp.	σ	B	$\rho_1(\hat{x})$	$\rho_2(\hat{x})$	$\rho_2(x^S)$	$\rho_3(\hat{x}, x^S)$
1	0.1	3	31.94	1.69	0.13	-1.57
2	0.3	3	6.65	0.23	0.15	-0.08
3	1	3	12.36	0.13	0.13	0.00
4	3	3	39.81	0.12	0.12	0.00
5	10	3	139.40	0.08	0.08	0.00
6	0.1	8	30.00	1.71	0.05	-1.64
7	0.3	8	12.08	0.46	0.17	-0.27
8	1	8	30.12	0.34	0.36	0.01
9	3	8	83.67	0.28	0.30	0.03
10	10	8	378.66	0.22	0.25	0.03

the same. Interestingly, for $\sigma = 10$ the value of $\rho_1(\hat{x})$ exceeds 100%, which shows that in this case the total variation error bound is reasonably tight and the MRP has difficulties proving this due to its statistical nature.

Remark 11. We can reduce bias or sampling error by increasing the sample size n or the sampling error by increasing N_r, and hence reduce $\rho_1(\hat{x})$ so that we do not obtain values above 100%. However, increasing n is computationally expensive. In particular, without using more sophisticated solution methods, it is very time consuming to solve the integer recourse problem in this section with a sample size larger than $n = 50$.

The results for multiplicative investment effects in Table 8 are similar. One difference is that for $b = 0.3$, the the solution of the convex approximation is also of high quality in the case of low variability. This explains why the values of $\rho_1(\hat{x})$ and $\rho_2(\hat{x})$ are low for $\sigma = 0.1$.

Table 8 Numerical results for the stochastic activity network investment problem. The effect of investments are multiplicative, $V^* = \{1, 6, 7, 12, 13\}$, and the ρ-values are reported as percentages. Positive values for ρ_3 mean that $\hat{x}$ outperforms x^S and negative values mean the opposite.

Exp.	σ	b	$\rho_1(\hat{x})$	$\rho_2(\hat{x})$	$\rho_2(x^S)$	$\rho_3(\hat{x}, x^S)$
1	0.1	0.3	0.00	0.00	0.03	0.02
2	0.3	0.3	0.20	0.007	0.03	0.02
3	1	0.3	2.00	0.02	0.05	0.03
4	3	0.3	53.11	0.20	0.23	0.03
5	10	0.3	280	0.20	0.25	0.05
6	0.1	0.7	29.48	1.85	0.07	-1.75
7	0.3	0.7	11.76	0.64	0.15	-0.48
8	1	0.7	5.90	0.17	0.20	0.01
9	3	0.7	30.50	0.27	0.32	0.02
10	10	0.7	105.55	0.18	0.17	0.01

We close this section with a brief discussion concerning trends for $\rho_1(\hat{x})$ with growing variance that we have seen throughout the paper and discussed in Section 4.3.2. The total variation bound appears in the denominator of $\rho_1(\hat{x})$, and as the variance of ω grows this error bound shrinks to zero. The numerator in $\rho_1(\hat{x})$ is the width of the MRP's confidence interval on the optimality gap. This width is governed by (i) suboptimality of the approximating solution, (ii) the bias in the lower bound estimator η_n^*, and (iii) the variance of $g(\hat{x},\omega) - g(x_n^*,\omega)$, where x_n^* solves (SP_n). For a fixed sample size, n, as the variance of ω grows we anticipate the contribution of (i) will shrink and the contributions of (ii) and (iii) will grow. The relative rates determine the behavior of $\rho_1(\hat{x})$. In most of our numerical experiments, starting at low variance, the shrinking of the numerator due to (i) dominates, and $\rho_1(\hat{x})$ initially shrinks as the variability of ω grows. Then when the variability of ω is sufficiently large, and the contribution of (i) is already small, the contributions of (ii) and (iii) grow and the denominator shrinks so that $\rho_1(\hat{x})$ grows. So, it is not surprising that when ω's variance is sufficiently large, $\rho_1(\hat{x})$ can exceed 100%. For example, when $\sigma = 10$ for experiment 10 in Table 7, the total variation error bound is $\rho_2(\hat{x})/\rho_1(\hat{x}) \times 100\% = 0.058\%$ of the optimal value. In principle, with this type of a priori guarantee from the error bound in hand there would be no reason to use the MRP to assess solution quality, although this is complicated by the fact that we obtain the approximating solution by solving a sample average approximation of the convex approximation.

6. Summary and conclusions

Two-stage integer recourse models can be very difficult to solve because they are non-convex. That is why we consider convex approximations for totally unimodular integer recourse models. In particular, we consider the α-approximations of van der Vlerk (2004) and the shifted LP-relaxation approximation developed in Romeijnders et al. (2016). Both approximations are obtained by simultaneously relaxing the second-stage integrality constraints and perturbing the distribution of the random right-hand side vector. The resulting approximating models can be considered as continuous recourse models, and can be solved efficiently by existing solution methods.

For both α-approximations and the shifted LP-relaxation approximation there are error bounds available that depend on the total variation of the probability density functions of the random variables in the model. The smaller these total variations, the smaller the error bounds, suggesting that the performance of the approximations is better in these cases. The actual performance, however, of these approximations had not yet been investigated.

We assess the quality of the approximating solutions using sampling. To do so, we use the multiple replications procedure of Mak et al. (1999), which can be used to assess the solution quality of a candidate solution for stochastic programming problems. We carry out numerical experiments on an integer newsvendor problem, a fleet allocation and routing problem, and an investment problem on a stochastic activity network. For these latter two problems we derive new error bounds to deal with the deterministic flow balance constraints in the second stage, and the fact that the recourse is relatively complete instead of complete; and to deal with perfect dependencies in the right-hand side random vector, respectively.

From these numerical experiments we conclude that the error bound is reasonably sharp if the variability of the random parameters in the model is either small or large; otherwise, the error bound is not so sharp, meaning that the actual error of using the convex approximation is much smaller than the error bound suggests. Moreover, we conclude that the performance of the convex approximation is good if the variability of the random parameters is medium to large. In case this variability is small, the performance of the approximations is not so good. However, these are precisely the cases where sampling methods perform well. In this sense, the convex approximations and sampling methods can be considered complementary solution methods for two-stage totally unimodular integer recourse problems.

Acknowledgments

The research of Ward Romeijnders is supported by The Netherlands Organisation for Scientific Research (NWO). The authors thank two anonymous referees whose suggestions improved this paper.

References

Ahmed, S., M. Tawarmalani, N.V. Sahinidis. 2004. A finite branch-and-bound algorithm for two-stage stochastic integer programs. *Mathematical Programming* **100** 355–377.

Avramdidis, A.N., K.W. Bauer, J.R. Wilson. 1991. Simulation of stochastic activity networks using path control variates. *Naval Research Logistics* **38** 183–201.

Bayraksan, G., D.P. Morton. 2006. Assessing solution quality in stochastic programs. *Mathematical Programming* **108** 495–514.

Bayraksan, G., D.P. Morton. 2009. Assessing solution quality in stochastic programs via sampling. M.R. Oskoorouchi, ed., *Tutorials in Operations Research*. INFORMS, 102–122.

Birge, J.R., F.V. Louveaux. 2011. *Introduction to Stochastic Programming*. Second edition. Springer, New York.

Carøe, C.C., R. Schultz. 1999. Dual decomposition in stochastic integer programming. *Operations Research Letters* **24** 37–45.

Donohue, C.J., J.R. Birge. 1995. An upper bound on the network recourse function. Working paper, Deartment of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan.

Elmaghraby, S.E. 1977. *Activity Networks: Project Planning and Control by Network Models*. Wiley, New York.

Freimer, M.B., J.T. Linderoth, D.J. Thomas. 2012. The impact of sampling methods on bias and variance in stochastic linear programs. *Computational Optimization and Applications* **51** 51–75.

Glynn, P.W., G. Infanger. 2013. Simulation-based confidence bounds for two-stage stochastic programs. *Mathematical Programming* **138** 15–42.

Higle, J.L., S. Sen. 1991. Stochastic decomposition: An algorithm for two stage linear programs with recourse. *Mathematics of Operations Research* **16** 650–669.

Higle, J.L., S. Sen. 1996. Duality and statistical tests of optimality for two stage stochastic programs. *Mathematical Programming* **75** 257–275.

Klein Haneveld, W.K., L. Stougie, M.H. van der Vlerk. 2006. Simple integer recourse models: convexity and convex approximations. *Mathematical Programming* **108** 435–473.

Klein Haneveld, W.K., M.H. van der Vlerk. 1999. Stochastic integer programming: general models and algorithms. *Annals of Operations Research* **85** 39–57.

Laporte, G., F.V. Louveaux. 1993. The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters* **13** 133–142.

Louveaux, F.V., R. Schultz. 2003. Stochastic integer programming. A. Ruszczyński, A. Shapiro, eds., *Handbooks in Operations Research and Management Science*, vol. 10: Stochastic programming. Elsevier, 213–266.

Mak, W.K., D.P. Morton, R.K. Wood. 1999. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters* **24** 47–56.

Prékopa, A. 1995. *Stochastic Programming*. Kluwer Academic Publishers, Dordrecht.

Rinnooy Kan, A.H.G., L. Stougie. 1988. Stochastic integer programming. Yu. Ermoliev, R.J-B Wets, eds., *Numerical techniques for stochastic optimization*, *Springer Ser. Comput. Math.*, vol. 10. Springer, 201–213.

Romeijnders, W., L. Stougie, M.H. van der Vlerk. 2014. Approximation in two-stage stochastic integer programming. *Surveys in Operations Research and Management Science* **19** 17–33.

Romeijnders, W., M.H. van der Vlerk, W.K. Klein Haneveld. 2015. Convex approximations of totally unimodular integer recourse models: A uniform error bound. *SIAM Journal on Optimization* **25** 130–158.

Romeijnders, W., M.H. van der Vlerk, W.K. Klein Haneveld. 2016. Total variation bounds on the expectation of periodic functions with applications to recourse approximations. *Mathematical Programming* **157** 3–46.

Ruszczyński, A. 1986. A regularized decomposition method for minimizing a sum of polyhedral functions. *Mathematical Programming* **35** 309–333.

Schrijver, A. 1986. *Theory of Linear and Integer Programming*. Wiley, Chichester.

Schultz, R. 2003. Stochastic programming with integer variables. *Mathematical Programming* **97** 285–309.

Sen, S. 2005. Algorithms for stochastic mixed-integer programming models. K. Aardal, G.L. Nemhauser, R. Weismantel, eds., *Handbooks in Operations Research and Management Science*, vol. 12: Discrete Optimization. Elsevier, 515–558.

Sen, S., J.L. Higle. 2005. The C^3 theorem and a D^2 algorithm for large scale stochastic mixed-integer programming: Set convexification. *Mathematical Programming* **104** 1–20.

Sen, S., Y. Liu. 2014. Mitigating uncertainty via compromise decisions in two-stage stochastic linear programming. Stochastic Programming E-Print Series.

Shapiro, A., D. Dentcheva, A. Ruszczyński. 2009. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia.

Shapiro, A., T. Homem-de-Mello. 1998. A simulation-based approach to two-stage stochastic programming with recourse. *Mathematical Programming* **81** 301–325.

van der Vlerk, M.H. 2004. Convex approximations for complete integer recourse models. *Mathematical Programming* **99** 297–310.

van der Vlerk, M.H. 2010. Convex approximations for a class of mixed-integer recourse models. *Annals of Operations Research* **177** 139–150.

van Slyke, R.M., R. Wets. 1969. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics* **17** 638–663.

Wallace, S.W., W.T. Ziemba, eds. 2005. *Applications of Stochastic Programming*. MPS-SIAM Series on Optimization.

Zverovich, V., C.I. Fábián, E.F.D. Ellison, G. Mitra. 2012. A computational study of a solver system for processing two-stage stochastic LPs with enhanced Benders decomposition. *Mathematical Programming Computation* **4** 211–238.